UNIVERSITY OF OULU

DEGREE PROGRAMME IN WIRELESS COMMUNICATIONS ENGINEERING

MASTER'S THESIS

ULTRA RELIABLE COMMUNICATION VIA OPTIMUM POWER ALLOCATION FOR REPETITION AND PARALLEL CODING IN FINITE BLOCK-LENGTH

Author _____
Endrit Dosti

Supervisor _____
Dr. Hirley Alves

Second examiner _____
Prof. Matti Latva-aho

Accepted _____ /_____ 2017

Grade _____

Dosti E. (2017) Ultra Reliable Communication via Optimum Power Allocation for Repetition and Parallel Coding in Finite Block-Length. Department of Communications Engineering, University of Oulu, Oulu, Finland. Master's thesis, 54 p.

ABSTRACT

In this thesis we evaluate the performance of several retransmission mechanisms with ultra-reliability constraints. First, we show that achieving a very low packet outage probability by using an open loop setup is a difficult task. Thus, we resort to retransmission schemes as a solution for achieving the required low outage probabilities for ultra reliable communication. We analyze three retransmission protocols, namely Type-1 Automatic Repeat Request (ARQ), Chase Combining Hybrid ARQ (CC-HARQ) and Incremental Redundancy (IR) HARQ. For these protocols, we develop optimal power allocation algorithms that would allow us to reach any outage probability target in the finite block-length regime. We formulate the power allocation problem as minimization of the average transmitted power under a given outage probability and maximum transmit power constraint. By utilizing the Karush-Kuhn-Tucker (KKT) conditions, we solve the optimal power allocation problem and provide closed form solutions. Next, we analyze the effect of implementing these protocols on the throughput of the system. We show that by using the proposed power allocation scheme we can minimize the loss of throughput that is caused from the retransmissions. Furthermore, we analyze the effect of the feedback delay length in our protocols.

Keywords: Ultra reliable communication, ARQ, CC-HARQ, IR-HARQ, Outage probability, Throughput, Delay, Finite blocklength.

Dosti E. (2017) Optimaalista tehoallokointia toisto- ja rinnakkaiskoodaukseen käyttävä erittäin luotettava tiedonsiirto äärellisillä lohkonpituuksilla.. Oulun yliopisto, sähkö- ja tietotekniikan osasto. Diplomityö, 54 s.

TIIVISTELMÄ

Tässä työssä arvioidaan usean uudelleenlähetysmenetelmän suorituskykyä erittäin luotettavan tietoliikenteen järjestelmäoletuksin. Aluksi osoitetaan, että hyvin alhaisen pakettilähetysten katkostodennäköisyyden saavuttaminen avoimen silmukan menetelmillä on haastava tehtävä. Niinpä työssä turvaudutaan uudelleenlähetyspohjaisiin ratkaisuihin, joilla on mahdollista päästä suuren luotettavuuden edellyttämiin hyvin alhaisiin katkostodennäköisyyksiin. Työssä analysoidaan kolmea uudelleenlähetysprotokollaa, nimittäin tyypin 1 automaattista uudelleen lähetystä (ARQ), Chase Combining -tyyppistä hybridi-ARQ -protokollaa (CC-HARQ) ja redundanssia lisäävää HARQ-protokollaa (IR-HARQ). Näille protokollille kehitetään optimaalisia tehon allokointialgoritmeja, joiden avulla päästään halutulle katkostodennäköisyystasolle äärellisillä lohkonpituuksilla. Tehon allokointiongelma muotoillaan keskimääräisen lähetystehon minimointiongelmaksi toteuttaen halutun katkostodennäköisyyden ja maksimilähetystehorajoituksen. Käyttämällä Karush-Kuhn-Tucker (KKT) -ehtoja ratkaistaan optimaalinen tehoallokointiongelma ja esitetään ratkaisut suljetussa muodossa. Seuraavaksi analysoidaan näiden protokollien järjestelmätason toteutusta läpäisykykytarkastelujen avulla. Niillä osoitetaan, että ehdotetulla tehon allokointimenetelmällä voidaan minimoida uudelleen lähetyksistä aiheutuvia suorituskykyhäviöitä. Lisäksi työssä tutkitaan takaisinkytkentäviiveen vaikutusta esitettyihin protokolliin.

Avainsanat: Erittäin luotettava tietoliikenne, ARQ, CC-HARQ, IR-HARQ, Katkostodennäköisyys, Läpäisykyky, Viive, Äärellinen lohkonpituus.

TABLE OF CONTENTS

FOREWORD

The focus of the thesis is in developing power allocation strategies that would enable reliable transmission of short packets. This research was carried out in the Center for Wireless Communications (CWC) as part of "5Gto10G" project. It was partially supported by Finnish Funding Agency for Technology and Innovation (TEKES), Huawei Technologies, Nokia and Anite Telecoms. I would like to thank Prof. Matti Latva-aho for giving me the opportunity to join his research group in CWC. I wish to express my gratitude to Dr. Hirley Alves for his guidance and numerous advices during the thesis. His vision and knowledge of the field have been very valuable during my work. Last but not least, I would like to thank my family for their support during my studies.

ABBREVIATIONS AND SYMBOLS

IoT	Internet of Things
4G	Fourth Generation of Mobile Communication Systems
5G	Fifth Generation of Mobile Communication Systems
MMC	Massive Machine-to-Machine Communication
MTC	Machine Type Communication
URC	Ultra Reliable Communications
ARQ	Automatic Repeat Request
ACK	Acknowledgment
NACK	Non Acknowledgment
HARQ	Hybrid Automatic Repeat Request
SNR	Signal-to-Noise Ratio
SC-HARQ	Selection Combining Hybrid Automatic Repeat Request
CC-HARQ	Chase Combining Hybrid Automatic Repeat Request
IR-HARQ	Incremental Redundancy Hybrid Automatic Repeat Request
FC IR-HARQ	Fixed Coding Incremental Redundancy Hybrid Automatic Repeat Request
VL IR-HARQ	Variable Length Incremental Redundancy Hybrid Automatic Repeat Reques
WiMAX	Worldwide Interoperability for Microwave Access
3GPP	Third Generation Partnership Project
LTE	Long-Term Evolution
MIMO	Multiple Input Multiple Output
AWGN	Additive White Gaussian Noise
LATR	Long Term Average Transmission Rate
ncpu	Nats per channel use
KKT	Karush-Kuhn-Tucker
EPA	Equal Power Allocation
Pr	Probability
log	Natural logarithmic function
min	Minimize function
M	Maximum number of transmissions
h	Channel gain
$\mathcal{CN}$	Normal distribution
z	Squared norm of the channel gain
e	Exponential function
m	One of the retransmission rounds
y_m	Received signal in the m^{th} round
ρ_m	The corresponding power term of a certain retransmission round
ρ_{avg}	Average transmit power
x_m	Transmitted signal in a certain retransmission round
w_m	AWGN term
N_0	Power spectral density of the noise
n	Number of channel uses
K	Number of information payload
ϵ	Target outage probability
$\mathcal{F}$	Encoder
c_i	One of the codewords

$\mathcal{G}$	Decoder
R	Channel coding rate
C_ϵ	Outage capacity
O()	Big O notation
C	Channel capacity
$Q()$	Gaussian Q-function
$E[\cdot]$	Expectation over the channel gain
V	Channel dispersion
k	Auxilliary variable
τ	Length of sub-codeword
∞	Infinity
E_M	Outage probability up to the final transmission round
ϵ_m	Outage probability of a certain transmission round
$\mathcal{L}$	Lagrangian function
μ	Inequality Lagrangian multiplier
λ	Equality Lagrangian multiplier
$g(R)$	Monotone and increasing function of the channel coding rate
$f(R)$	Monotone and increasing function of the channel coding rate
H	Hessian matrix
$\mathcal{R}$	Set of real numbers

1. INTRODUCTION

Mobile communication systems play a very important role in everyone's daily life. It would be very difficult to imagine modern life without wireless communication, as it existed before 1990. Throughout the history of the development of these systems, there have been many changes. First and second generation mobile communications were dominated by analog and then digital audio signals and text messaging. The third generation was more about scaling the numbers of users on the network for voice communications and text messaging but was overwhelmed by an unpredictable tsunami of image and video content [1]. Next, to cope with the increasing demand for higher data rates, the fourth generation of mobile communications (4G) emerged. Mobile 4G Long Term Evolution (LTE) was the first global standard for mobile broadband communications, and managed to satisfy the data rate requirements for most of the applications that were in the market at the time [2].

The trend of providing higher data rates is expected to continue in the next generations of mobile communication systems as well. However, in these systems, there will be several new features. The vision will be to connect all devices that benefit from an internet connection [3], thus resulting in the creation of the Internet of Things (IoT). A key characteristic of the IoT is that most of the wireless connections are expected to be generated by the autonomous devices rather than by the human-operated terminals. To successfully implement this vision, wireless communication systems will have to support a much larger number of connected devices and at the same time fulfill much more stringent requirements on latency and reliability that what current standards can guarantee [4].

For this purpose, in the fifth generation of mobile communication systems (5G), at least two new operating modes will be introduced. The first mode, which is Massive Machine-to-Machine Communication (MMC), is related to designing a wireless system that can support a large number of devices (e.g. more than 10 000) connected at the same time. This will pave path for seamlessly and ubiquitous connectivity foreseen in 5G and IoT [5, 6]. The massive Machine Type Communication (mMTC) networks have a massive number of devices communicating with diverse range of requirements in terms of reliability, latency, data rates, and energy consumption, besides diverse traffic patterns [7]. Traditionally, machine type connectivity has relied on short-range wireless technologies. However, moving towards large-scale deployments requires broader interconnection capabilities which are best enabled by the wide-area coverage of cellular network infrastructures, especially with the current research, development and standardization efforts towards a global 5G network [8].

Massive MTC covers a diverse range of new services and applications. They can be utilized in metering services, such as electricity, gas and water metering. These kind of services will require the support of a very large number of connected devices (very high density), which generate infrequent and small amount of data [9]. Another class of services are control and monitoring systems. They play a very important role in real-time control and automation of industrial or home environments. As such, they come with very stringent requirements on high reliability and low-latency transmissions [10]. Furthermore, mMTC can be utilized in tracking applications. They can assist in fleet management, prevention of theft of equipments and asset tracking. Supporting these kind of services requires large-scale connectivity and low power consumption. More-

over, they can be utilized in making payments in different points of sale or vending machines. For these kind of applications, security is of vast importance. Finally, mMTC systems play a very important role in security and public safety services. Typical applications in which these systems can be deployed include surveillance systems, home security and access control. The key requirements for their successful deployment are high security, reliability and low latency [11].

The second operating mode will be Ultra Reliable Communications (URC). This mode comprises of two basic functions, reliability and latency. Reliability refers to the capability of guaranteeing successful transmissions (i.e. very low outage probability) within a certain delay budget. The requirements for it are far more stringent than in the case of mMTC and may vary among different applications. Often, the outage probability requirement can go as low as 10^{-9}. The other function, latency, refers to the time delay between the moment when the data is generated, until the moment when this data is correctly received. Its requirements can again vary between different applications, but they can be lower than 1 ms [12, 13, 14]. In conclusion, URC refers to insuring a certain level of communication service almost 100% of the time, while satisfying very stringent delay requirements (e.g. less than 1 ms) [15]. It is not present in traditional wireless networks, as they have not been designed to meet such requirements. For example, the LTE network can achieve error rates within the range of 10^{-2} to 10^{-1} when the end-to-end latency is on the order of few miliseconds [16].

Since it is an entirely new feature, modern systems will need to be designed in order to support URC. It is expected to be very important in many applications, such as reliable wireless coordination among vehicles. In future systems, vehicles will be exchanging information wirelessly. Through short messages, they will be able to communicate with each-other, traffic lights and road signals (cooperative intelligent transport systems). Making such systems, will require the design of completely new transmission techniques and radio access protocols [17]. Another important application of URC is on reliable cloud connectivity. In the cloud-based systems that we have nowadays, internet connectivity is presumed to be available most of the time. Furthermore, the way they are designed is entirely based on the reliability of the connectivity. Therefore, as the connection becomes more reliable, the way these cloud services are offered will be reshaped. For example, they can be designed to provide a certain service 99.9999% of the time with low data rates (e.g. 0.5 Mbps) while maintaining low latency (e.g. 0.5 ms); or they can provide service 95% of the time, with moderate data rates (e.g. 70 Mbps) and latency (e.g. 5 ms) [15]. Finally, URC is expected to play a very important role in critical connections for industrial automation. A good illustration of this are modern factories, which are relying on automatic guided vehicles, unmanned shuttles and mobile equipments. This results in increased efficiency. Furthermore, the utilization of wireless communication in these environments can also help with avoiding wiring and space constraints for control and monitoring processes [18]. However in these scenarios, wireless links will be carrying both redundant and vital information, which is essential to be delivered correctly. Designing such systems can be quite challenging, especially when considering that they will have to co-exist with usual low-band traffic [19, 20].

When designing wireless communication systems, one of the biggest challenges is mitigating the presence of fading. This phenomena is mainly caused by multipath propagation and several time-varying effects (e.g. environment changes, mobility be-

tween transmitter and receiver etc). It causes fluctuations in the received signal power, thus resulting in loss of transmitted packets [21]. To cope with fading, several schemes or processing algorithms have been implemented in the transmitter or receiver. They are known as diversity schemes and their goal is to create signal redundancy, which is typically done by transmitting the signal via several independent fading paths and then combining everything at the receiver.

In this context, the utilization of different retransmission schemes is classical approach to provide larger diversity gains. The most popular are repetitive and parallel retransmission schemes. In the first scheme, the transmitter sends the same codewords in all the possible fading paths. Then, based on the communication protocol that is implemented, the receiver tries to recover the information. In the second scheme, the transmitter utilizes different and jointly designed codewords to construct the packets. These packets, which will contain different information, are then combined at the receiver.

These schemes are embraced by several retransmission protocols. The simplest retransmission protocol which utilizes repetition coding, is Automatic Repeat Request (ARQ) [22, 23]. In this protocol, the transmitter sends the packets until it receives positive confirmation (ACK) from the receiver or until the maximum allowed number of retransmissions is exhausted. Another retransmission protocol family are Hybrid-ARQ (HARQ) protocols [24]. These protocols are similar to the previous ones, in the sense that they utilize repetition channel coding as well. However, the difference is that the receiver buffers all the packets and utilizes them to correctly decode the information. Many HARQ protocols exist and are classified based on how the receiver combines the packets. For instance, when the receiver selects the packet with the highest Signal-to-Noise Ratio (SNR) and makes decisions based on it, the scheme is called Selection Combining (SC) HARQ [25]. Secondly, the receiver can do Maximal Ratio Combining (MRC) of the received signals and extract the information. This scheme is known as Chace Combining (CC) HARQ [26, 27, 28]. Both schemes highly increase the probability of having successful transmissions. However, they can cause significant losses in the system throughput. To minimize the losses, several retransmission protocols utilize parallel coding schemes. These protocols utilize Incremental Redundancy (IR) schemes [29, 30, 31]. Herein, the transmitter sends new information with each retransmission. This is achieved by splitting the parent codeword into several sub-codewords and sending them. Depending on the length of the sub-codewords, there are two main schemes. If all the sub-codewords have same length, then the scheme is Fixed Coding (FC) IR-HARQ. If the sub-codewords have different lengths, then the scheme is called Variable Coding (VC) IR-HARQ. The utilization of this family of retransmission protocols is a well-established approach, which has been embraced by several systems, such as Worldwide Interoperability for Microwave Access (WiMAX), Third Generation Partnership Project Long-Term Evolution (3GPP LTE), 3GPP LTE-IOT [32, 33, 34]

The problem of allocating different power levels in each round of a retransmission scheme has been investigated in several papers [35, 36, 37]. In [37, 38] the authors formulate a geometrical programming problem (GPP), which they utilize to develop a power allocation scheme for HARQ protocol when the maximum number of transmissions is set to two. Furthermore, they provide a closed form approximation of the outage probability. In [38], the authors prove the equivalence of two optimization

problems for the case of IR-HARQ protocol. Again, they provide a power allocation strategy only for the case when the maximum number of allowable transmissions in set to two. The root finding solution that is proposed therein, computes all the saddle points and selects the smallest one. When the number of saddle points is small, that approach can give a reasonably good performance. However, as the number of saddle points increases, computing the optimal power terms, becomes more complex. Additionally, in none of those papers, the authors do not provide any closed form solutions for the general case of M transmissions. In [35] the authors provide asymptotic lower and upper bounds for the optimal transmission rate of a Multiple Input Multiple Output (MIMO) ARQ system. However, they do not take into consideration any power allocation strategies.

Throughout all the papers that were discussed above, the analysis is done under the assumption of asymptotically long codewords. This implies that the length of metadata (control information) is much smaller than the actual data, thus its effect can be neglected in the calculations. In the finite block-length regime this assumption does not hold anymore. Metadata and the actual data are almost of the same size, therefore the usage of conventional methods, such as capacity or ergodic capacity, is highly suboptimal [4, 39]. Thus, the performance of these schemes with power allocation should be evaluated. Little work has been done in this field for the short packets domain. For instance, [40] evaluates the performance of IR-HARQ for aditive white Gaussian noise (AWGN) channel showing that a large number of retransmissions enhances performance in terms of the long term average transmission rate (LATR). However, authors do not assess the impact of increasing the number of retransmissions on latency, nor do they guarantee high reliability. Further, in [41] the authors analyze the performance of ARQ protocol over the fading channel under very simplistic assumptions such as infinite number of transmissions, full buffer capacity, instantaneous and error free feedback. Therein, authors do not investigate the impact of power allocation between different ARQ rounds. In [42], the authors develop a power allocation scheme for type-I ARQ protocol that minimizes the outage probability only for the case of two transmissions. However, in their scheme they do not guarantee a minimal outage probability level which would be essential in the case of URC, since different applications have different reliability requirements.

1.1. Thesis contributions

In this thesis, we develop power allocation algorithms for different repetitive and parallel retransmission protocols, that would allow us to achieve any reliability requirement in the finite blocklength regime. First, we prove that it would not be feasible to operate in the ultra reliable region by having only one transmission of the packet (open loop setup). For this reason, we analyze different retransmission protocols. We formulate the power allocation problem as an optimization problem and prove that it is convex. The problem is solved for each of the protocols by using Karush-Kuhn-Tucker (KKT) conditions for any number of retransmissions. Furthermore, we suggest a simpler way of solving the problem, which is valid for the case when the maximum number of transmissions is set to two.

The main contributions of the thesis are as follows

- Power allocation algorithms are developed for *i)* Type I ARQ; *ii)* CC-HARQ; and *iii)* IR-HARQ.
- We provide a closed-form approximation for the outage probability of the IR-HARQ protocol when variable power is allocated in each round.
- We compare the power efficiency of the algorithms that we have developed.
- We show that the proposed power allocation strategies minimize the effect of retransmissions in the overall system throughput.
- We evaluate the effect of retransmissions in the overall system delay and provide comparisons between the different protocols.

Finally, from the work done for this thesis, we have managed to write the following papers:

- E. Dosti, U. L. Wijewardhana, H. Alves, and M. Latva-aho, "Ultra reliable communication via optimum power allocation for Type-I ARQ in finite Block-Length," in IEEE ICC 2017 Wireless Communications Symposium (ICC'17 WCS), Paris, France, May 2017, pp. 5019-5024.
- E. Dosti, M. Shehab, H. Alves, and M. Latva-aho, "Ultra reliable communication via CC-HARQ in finite Block-Length," in 2017 European Conference on Networks and Communications (EuCNC): Physical Layer and Fundamentals (PHY) (EuCNC2017 - PHY), Oulu, Finland, Jun. 2017.
- E. Dosti, H. Alves, and M. Latva-aho, "Ultra reliable communication via optimum power allocation for repetitive and parallel retransmission schemes in finite Block-Length," to be submitted to IEEE Transactions on Wireless Communications.

1.2. Thesis outline

The rest of the thesis is organized as follows. Chapter 2 introduces the system model and defines the communication at finite block-length. Next, in Chapter 3 the optimization problem is formulated and the power allocation algorithms for the three different retransmission protocols are developed. Then, Chapter 4 analyzes the throughput and delay performance of our algorithms. Chapter 5 provides the numerical results of the work. Finally, Chapter 6 provides the conclusions and suggest future work.

2. MAXIMUM CODING RATE IN FINITE BLOCK-LENGTH

The focus of the thesis is the evaluation of different retransmission in a URC setup. In this chapter we first introduce the system model. We show the gap between the channel capacity and the maximum achievable rate. Then, we utilize the most recent results from information theory to characterize the communication at finite block-length. Furthermore, we show that when the transmission SNR is high enough, utilizing the first order Taylor expansion of the asymptotic approximation gives the same performance as the closed form approximation or integral solution. Moreover, we show that it would not be feasible to utilize an open loop setup to convey the information under stringent reliability requirements. Numerically, we show that to meet such requirements with only one transmission of the packet, we would be needing to transmit either a very large amount of power, or with very low channel coding rate.

2.1. System model

Assume a transmitter-receiver pair communicating with each-other. In the transmitter side, $b_1, b_2, \ldots, b_K$ nats[1] are encoded in $c_1, c_2, \ldots, c_{n_c}$. Next, these encoded nats are interleaved and mapped to a constellation $\mathcal{X}$. This results in the stream of modulated symbols $x_1, x_2, \ldots, x_n$. For simplicity, we assume that we map one modulated symbol per channel use. Here, K and n denote the number of information nats and the number of channel uses, respectively. The scheme of the transmitter is shown in in Figure. 2.1 The receiver then fetches the packets and tries to recover the information. First, it performs demodulation. Next, the demodulated symbols are de-interleaved and decoded. Afterwards, the stream of nats is fed to the next blocks for further processing.

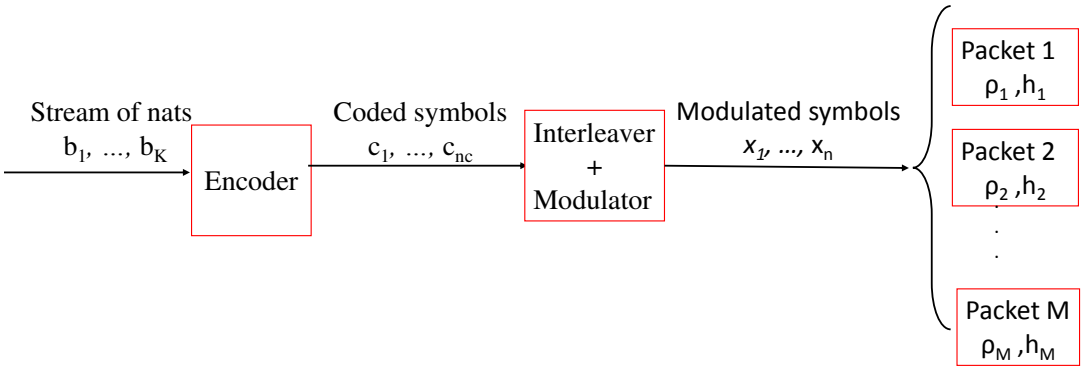


Figure 2.1: Structure of transmitter. First, the stream of information nats is encoded. Then, the encoded symbols are interleaved and mapped to a pre-defined constellation. This results in the creation of the packets that will be transmitted.

While communicating, the pair can either utilize an open loop setup, or a retransmission protocol. In the second case, the maximum number of transmissions is set

[1]Notice that hereafter in order to standardize the notation we assume that all information is encoded in nats instead of bits. Therefore, all $\log$ is the natural logarithm.

to M. Whenever the transmitter fetches a non acknowledgment (NACK) packet, it retransmits based on the protocol that is implemented. It will stop sending packets if it receives an acknowledgment packet (ACK) from the receiver, or if the maximum number of allowed transmissions has been completed.

We consider quasi-static fading channel conditions, in which the channel gain h remains constant for the duration of one packet transmission and changes independently between all the transmission rounds. We analyze the case when the channel coefficient is Rayleigh distributed and $h \sim \mathcal{CN}(0,1)$. Thus the squared envelope of the channel gain is exponentially distributed with mean one. For simplicity we denote $f_{|h|^2}(z) = e^{-z}$. We assume that the receiver has channel state information while the transmitter knows only the distribution of the channel gains and the information it obtains from the feedback. Then the received signal at the m^{th} round can be written as

$$y_m = \sqrt{\rho_m} h_m x_m + w_m, \tag{1}$$

where x_m is the transmitted signal and w_m is the AWGN noise term with noise power $N_0 = 1$. The term ρ_m is the packet transmit power, which since the variance of the noise is set to 1, corresponds to the transmission signal-to-noise ratio (SNR).

2.2. Communication at finite block-length

In this section, we briefly summarize the recent results in the characterization of the maximum channel coding rate and outage probability in the finite block-length regime. Further, we evaluate the case of the open loop setup.

As in [43], for notational convenience we need to define an (n, K, ρ, ϵ) code as a collection of

- An encoder $\mathcal{F} : \{1, \ldots, K\} \mapsto \mathcal{C}^n$ which maps the message $k \in \{1, \ldots, K\}$ into an n-length codeword $c_i \in \{c_1, \ldots, c_n\}$ such that the following power constraint (ρ) is satisfied:

$$\frac{1}{n}\|c_i\|^2 \leqslant \rho, \forall i. \tag{2}$$

- A decoder $\mathcal{G} : \mathcal{C}^n \mapsto \{1, \ldots, K\}$ that satisfies the maximum error probability (ϵ) constraint:

$$\max_{\forall i} \Pr\left[\mathcal{G}(y) \neq I | I = i\right] \leqslant \epsilon, \tag{3}$$

  where y is the channel output induced by the transmitted codeword according to (1).

The maximum achievable rate of the code is defined as:

$$R^*_{max}(n, \rho, \epsilon) = \sup\left\{\frac{\log K}{n} : \exists (n, K, \rho, \epsilon) \text{ code}\right\}, \tag{4}$$

where $\log$ refers to the natural logarithm.

For the AWGN channel non-asymptotic lower and upper bounds on the maximum achievable rate have been derived in [44]. Based on their analysis, when the block-length is large enough, the maximum coding rate in nats can be found as

$$R^*_{max}(n,\rho,\epsilon) = C - \sqrt{\frac{V}{n}}Q^{-1}(\epsilon) + \frac{\log n}{2n} + O(1), \tag{5}$$

where $Q(\cdot)$ denotes the Gaussian Q-function, $O(\cdot)$ denotes the remainder term, C denotes the channel capacity and the channel dispersion V. They are also computed in ncpu and can be found as

$$C(\rho) = \log(1+\rho), \tag{6}$$

$$V(\rho) = 1 - \frac{1}{(1+\rho)^2}. \tag{7}$$

From (5) we notice that there is a penalty which is assigned to the maximum achievable rate R_{max*} as a consequence of small block sizes. This penalty is represented from the channel dispersion and a remainder term. Furthermore, we observe that R^*_{max} increases as ϵ increases (this is obvious since the inverse Q-function $Q^{-1}(\epsilon)$ is a decreasing function of the outage probability). This implies that, we can achieve transmission with arbitrarily high channel coding rate, as long as we allow higher error probability. Also, from (7) we notice that as the transmission power ρ or the size of the block-length increase, the channel dispersion goes to zero.

Another important result obtained from (5), is that for sufficiently large n (e.g. greater than 50 [41]) the minimum packet error rate encountered in the transmission of K nats over the AWGN channel by utilizing packets of n channel uses can be well approximated by

$$\epsilon(n,R,\rho) \approx Q\left(\frac{nC(\rho) + 0.5\log n - K}{\sqrt{nV(\rho)}}\right), \tag{8}$$

Finally, from (5) we notice that as $n \to \infty$ the terms $\frac{\log n}{n}$ and $O(1)$ vanish. Therefore, we can deduct that as the transmission SNR or the number of channel uses increase, the R_{max} will converge to the channel capacity.

In Figure 2.2 we analyze the behavior of the maximum coding rate as a function of the size of the packet that will be transmitter (in channel uses). We plot for different values of transmission SNR, while fixing the outage probability to $\epsilon = 10^{-3}$. From the plot we observe the gap between the channel capacity and maximum achievable rate. Further, we notice that this gap becomes smaller as ρ and n increase, a result which is coherent with the mathematical observations that we made above.

Recently, a tight approximation for $R^*_{max}(n,\rho,\epsilon)$ has been proposed for sufficiently large values of n in the case of the quasi-static fading channel [43] and is given by

$$R^*_{max}(n,\rho,\epsilon) \approx C_\epsilon + \mathrm{O}\left(\frac{\log n}{n}\right), \tag{9}$$

where C_ϵ is the outage capacity:

$$C_\epsilon = \sup\{R : \Pr[\log(1+\rho\cdot z) < R] < \epsilon\}. \tag{10}$$

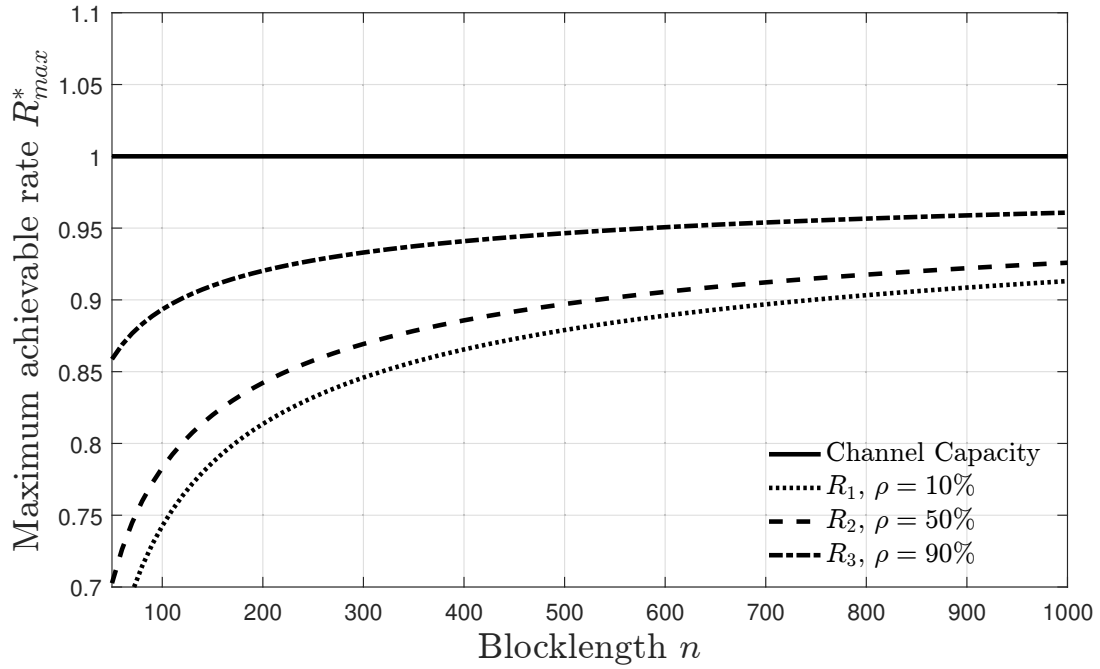


Figure 2.2: Maximum achievable rate as a function of the number of channel uses for different values of transmission SNR.

Then, by assuming a code rate of $R = \frac{K}{n}$ nats per channel use (ncpu), where K is the information payload, the outage probability is approximated as [45]

$$\epsilon(n, R, \rho) \approx E\left[Q\left(\frac{C(\rho \cdot z) - R}{\sqrt{V(\rho \cdot z)}}\right)\right] \tag{11}$$

$$\approx \int_0^\infty e^{-z} Q\left(\frac{C(\rho \cdot z) - R}{\sqrt{V(\rho \cdot z)}}\right) \mathrm{d}z, \tag{12}$$

where $E[\cdot]$ denotes the expectation over the channel gain z. However the integral in (11) does not have a closed form solution. Thus, we resort to an approximated closed-form expression as in [46]

$$\epsilon(n, R, \rho) = 1 - \frac{\delta}{\sqrt{2\pi}} e^{-\kappa} \left(e^{\sqrt{\frac{\pi}{2\delta^2}}} - e^{-\sqrt{\frac{\pi}{2\delta^2}}}\right), \tag{13}$$

where κ and δ are given from (14) and (15)

$$\kappa = \frac{e^R - 1}{\rho} \tag{14}$$

$$\delta = \sqrt{\frac{n\rho^2}{e^{2R} - 1}} \tag{15}$$

Note that (13) characterizes the outage probability of a single ARQ round.

Fig. 2.3 illustrates the outage probability for the open loop setup, where the message is conveyed in a single transmission, for different channel coding rates. We

have fixed the number of channel uses $n = 200$ and analyzed the case of mapping $K \in \{200, 400, 600\}$ information nats. This results in the channel coding rates $R = 1$, $R = 2$ and $R = 3$ ncpu, respectively. We can see that the integral form in (8) and the closed-form approximation in (13) match well for all the coding rates.

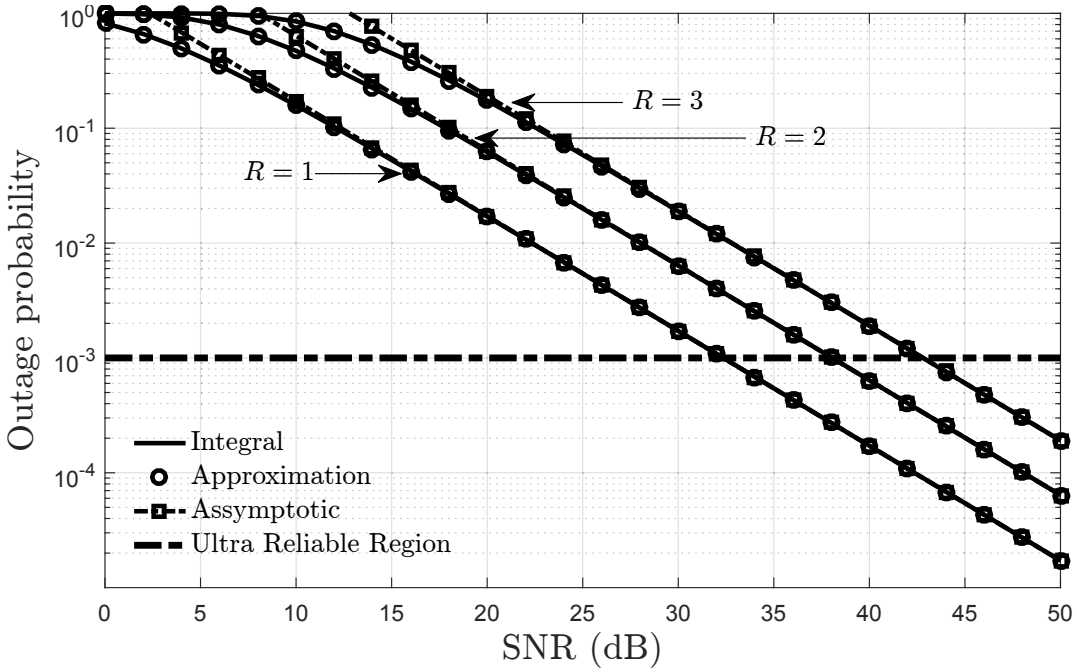


Figure 2.3: Outage probability for the open loop setup for different channel coding rates.

Furthermore, in Fig. 2.3 we also illustrate the performance of the asymptotic approximation (which we derive next and is given by (17)). The results show that in the ultra reliable regime, where very low outages are required $\epsilon < 10^{-3}$, the asymptotic approximation can be used. This is due to the fact that in high SNR the maximum achievable rate (9) converges to the one with asymptotically long codewords $R^*(n, \epsilon) = C_\epsilon$, where C_ϵ is defined in (10). Therefore, when $\rho \to \infty$ the outage probability in the m^{th} round can be calculated as:

$$\epsilon_m = \Pr[\log(1 + \rho \cdot z) < R] = 1 - e^{\frac{e^R - 1}{\rho_m}}. \tag{16}$$

Note that (16) holds for Rayleigh fading channels. Furthermore, for (16) we can utilize the Taylor expansion. Since we are focusing our analysis in the high SNR regime, the entire Taylor series can be well-approximated from the first order of Taylor expansion $e^{-x} \approx 1 - x$. This way, we can express the asymptotic approximation for the outage probability of the m^{th} round as:

$$\epsilon_m = \frac{\phi}{\rho_m}, \tag{17}$$

where $\phi = e^R - 1$.

To guarantee URC, we require to have an outage probability ϵ very low. However, Figure 2.3 shows that such low outage values are highly unlikely to be obtained when

there is a suboptimal allocation of the number of information nats K, number of channel uses n or the transmission SNR ρ. For this purpose, we investigate numerically what would be the optimal number of information nats or channel uses that we need to allocate.

First, in Figure 2.4 we analyze the behavior of the outage probability as a function of K. We set the transmission SNR $\rho \in \{20, 30\}$ dB, which are values that are high enough to ensure operation in the ultra reliable region where the assymptotic approximation fully matches the closed-form approximation (as shown from Figure 2.3). Furthermore, we change the transmission rate by utilizing different values for the number of channel uses $n \in \{100, 200, 300\}$. From the results obtained, we observe that the optimisation problem is not strictly convex. Specifically, we notice that increasing the transmit power, while maintaining transmission at low channel coding rate ($R = \frac{1}{3}$ ncpu), allows us to achieve low outage values. However, these outage values are far from the targets that would like to have when discussing about URC. Furthermore, it would not be feasible to convey such low amount of information, and at the same time spend very large amounts of power.

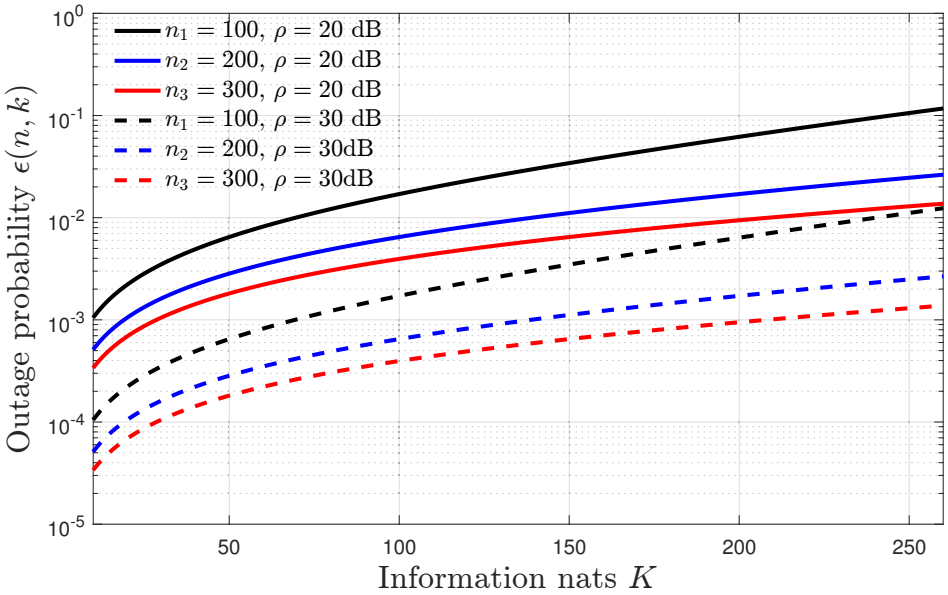


Figure 2.4: Outage probability versus the number of information nats for different values of transmission SNR and channel uses.

In practical systems, the amount of information that a node has to convey is in general fixed. Therefore, it would be of high interest to evaluate the optimal number of channel uses. For this purpose, we fix again the transmission SNR $\rho \in \{40, 50\}$ dB. Next, we set the number of information nats to be conveyed as $K \in \{100, 200, 300\}$. The results are shown in Figure 2.5. Similarily as above, we notice that to achieve very low outage probability we need to transmit at very high power levels, and at the same time maintain low channel coding rates.

From the plots shown above, respectively Figures 2.3, 2.4 and 2.5 we can deduct that it would be impossible to obtain the targeted outage probability values while utilizing the open loop setup. In order to obtain an outage probability in the ultra reliable region,

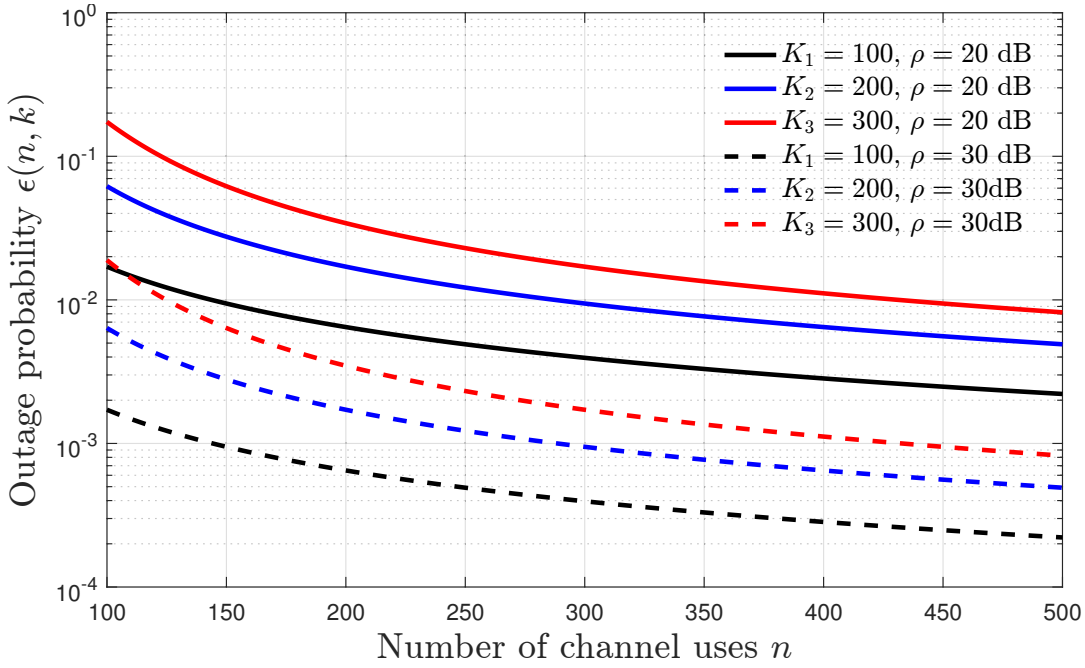


Figure 2.5: Outage probability versus the number of channel uses for different values of transmission SNR and channel uses.

and at the same time spend as little resources as possible, we investigate the possibility of utilizing retransmission mechanisms with optimal power allocation.

3. POWER ALLOCATION

In this chapter we evaluate the impact of repetitive and parallel retransmission schemes in the outage probability. We propose optimal power allocation algorithms that allow us to reach any target outage probability assuming that we can have up to M-transmissions for each of the protocols. Furthermore, we propose a simpler method of obtaining the power terms in the case when the maximum number of transmissions is fixed to $M = 2$.

The problem of interest is to achieve a target outage probability while spending as little power as possible for sending the information from the transmitter to the receiver. Since, we will have multiple transmissions, one approach would be to allocate equal power in each round. This implies that given a certain power budget ρ_{budget}, the transmit power in the m^{th} round would be $\rho_m = \frac{\rho_{budget}}{M}$. However, such simplistic approach is shown to be highly inefficient for very low outage probability values for different retransmission schemes [47]. Thus, we propose a power allocation algorithm in order to minimize the average transmit power of the transmitter which allocates different power levels in each retransmission round. Bearing this in mind, the average transmitted power can be defined as

$$\rho_{avg} = \frac{1}{M} \sum_{m=1}^{M} \rho_m E_{m-1}, \tag{18}$$

where M is the maximum number of retransmission rounds, ρ_m is the power transmitted in the m^{th} round and E_{m-1} is the outage probability up to the $m-1$ round. Next, we calculate the probability that the packet can not be decoded correctly even after the maximum allowed number of retransmissions. We refer to this as packet drop probability (pdp), and it corresponds to the outage probability up to the M^{th} round (E_M). Since we have assumed that all the transmissions of the packets experience independent fading conditions we can express E_M as

$$E_M = \prod_{m=1}^{M} \epsilon_m, \tag{19}$$

where ϵ_m is the outage probability of the m^{th} transmission round. The outage probability before the first transmission, $\epsilon_0 = 1$.

Now, we can formulate the following power allocation problem:

$$\begin{aligned} \text{minimize} \quad & \rho_{\text{avg}} \\ \text{subject to} \quad & 0 \leqslant \rho_m, \ 1 \leqslant m \leqslant M \\ & E_M = \epsilon \end{aligned} \tag{20}$$

where ϵ is any target outage probability. Since the optimization problem (20) is nonlinear and the feasible set is compact, we can find a global optimal solution [38]. Furthermore, the problem is convex, and the proof is shown in Appendix 7.2. Therefore, we can utilize the Karush-Kuhn-Tucker (KKT) conditions to obtain the optimal solution for the convex problem (20).

3.1. Repetitive retransmission schemes

The utilization of repetitive retransmission schemes implies that in each round the transmitter sends the same information nats. In this section we present the power allocation algorithms for two retransmission protocols that utilize repetitive schemes. Furthermore, we discuss about the main differences between them.

3.1.1. Type-I ARQ

The first member of repetitive schemes that we analyze is Type-I ARQ protocol. Its principle is shown in Figure. 3.1. When utilizing this scheme, each transmission is

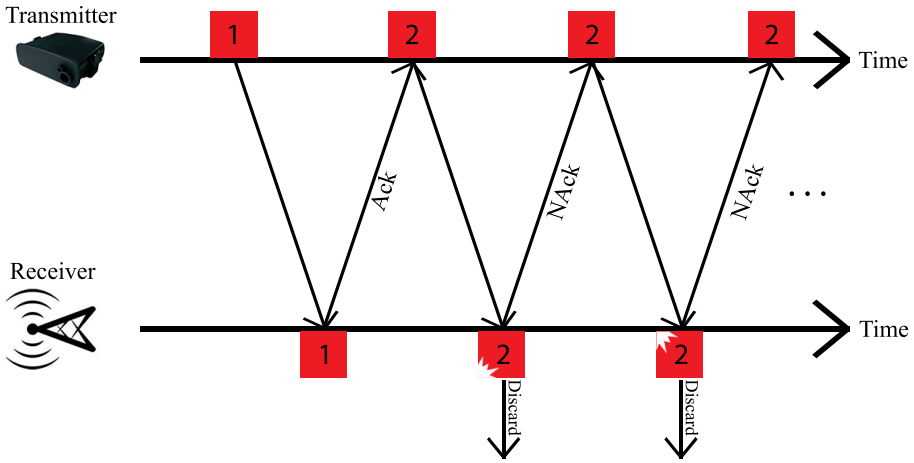


Figure 3.1: The setup for Type-I ARQ protocol. The transmitter sends first packet 1 and waits for the Ack from the receiver. Then it sends packet 2. If the receiver can not decode the packet, it discards the packet and asks for retransmission.

done with full rate $R = \frac{K}{n}$. This implies that, the transmitter sends the entire packet in each transmission round and stops when the maximum number of transmissions M has been achieved, or when it receives confirmation that the packet has been successfully decoded. The receiver makes the decisions only based on the last packet he has received, and discards the earlier packets.

The outage probability for the m^{th} ARQ round and can be computed by (13), or asymptotically via (17). We start the solution for problem (20) by writing the Lagrangian function, which is computed as

$$\mathcal{L}(\rho_m, \mu_m, \lambda) = \frac{1}{M} \sum_{m=1}^{M} \rho_m E_{m-1}^{ARQ} + \sum_{m=1}^{M} \mu_m \rho_m + \lambda (E_M^{ARQ} - \epsilon) \tag{21}$$

where μ_m for $m = 1, \ldots, M$ and λ are the Lagrangian multipliers. Now, we express the KKT conditions as follows:

C1 $\frac{\partial \mathcal{L}}{\partial \rho_m} = 0,\ m = 1, \ldots, M,$

C2 $\mu_m \geqslant 0,\ m = 1, \ldots, M,$

C3 $\mu_m \rho_m = 0,\ m = 1, \ldots, M,$

C4 $E_M^{ARQ} - \epsilon = 0.$

We can write the derivative of the Lagrangian function $\mathcal{L}(\rho_m, \mu_m, \lambda)$ with respect to the power ρ_m as

$$\frac{\partial \mathcal{L}(\rho_m, \mu_m, \lambda)}{\partial \rho_m} = \frac{1}{M}\left(\frac{\phi^{m-1}}{\prod_{i=1}^{m-1}\rho_i} - \sum_{i=1}^{M-m}\frac{\rho_{m+i}\phi^{m+i-1}}{\rho_m^2\prod_{j=1,j\neq m}^{m+i-1}\rho_j}\right) - \mu_m - \lambda\left(\frac{\phi^M}{\rho_m^2\prod_{i=1,i\neq m}^{M}\rho_i}\right). \tag{22}$$

The transmit power at m^{th} ARQ round $\rho_m > 0$ for $m = 1, \ldots, M$ since we require some power to transmit the information at each of these rounds. Thus, from the complementary slackness condition (**C3**) we can see that $\mu_m = 0$ for $m = 1, \ldots, M$. Now, using (22) and $\mu_M = 0$, we can write **C3** for $m = M$ as

$$\frac{\partial \mathcal{L}(\rho_m, \mu_m, \lambda)}{\partial \rho_M} = \frac{1}{M}\frac{\phi^{M-1}}{\prod_{i=1}^{M-1}\rho_i} - \lambda\left(\frac{\phi^M}{\rho_M^2\prod_{i=1,i\neq m}^{M}\rho_m}\right) = 0. \tag{23}$$

After some algebraic manipulations of (23) we obtain the transmit power at the M^{th} ARQ round as

$$\rho_M = \sqrt{M\lambda\phi}. \tag{24}$$

Similarily substituting $m = M-1$ in (22) and using $\mu_{M-1} = 0$, we can rewrite **C1** for $m = M-1$ as

$$\frac{\partial \mathcal{L}(\rho_m, \mu_m, \lambda)}{\partial \rho_{M-1}} = \frac{1}{M}\left(\frac{\phi^{M-2}}{\prod_{i=1}^{M-2}\rho_i} - \frac{\rho_M\phi^{M-1}}{\rho_{M-1}^2\prod_{i=1}^{M-2}\rho_i}\right) - \lambda\left(\frac{\phi^M}{\rho_{M-1}^2\prod_{i=1,i\neq M-1}^{M}\rho_i}\right) = 0. \tag{25}$$

Mathematical simplification of (19) leads to

$$\rho_{M-1} = \sqrt{\rho_M\phi + \frac{M\lambda\phi^2}{\rho_M}}. \tag{26}$$

Following a similar approach, we can write the condition **C1** for the case of $m = M-2$ and obtain

$$\frac{\partial \mathcal{L}(\rho_m, \mu_m, \lambda)}{\partial \rho_{M-2}} = \frac{1}{M}\left(\frac{\phi^{M-3}}{\prod_{i=1}^{M-3}\rho_i} - \sum_{i=1}^{2}\frac{\rho_{M-2+i}\phi^{M-3+i}}{\rho_{M-2}^2\prod_{j=1,j\neq M-2}^{M-3+i}\rho_j}\right) - \lambda\left(\frac{\phi^M}{\rho_{M-2}^2\prod_{i=1,i\neq M-2}^{M}\rho_i}\right). \tag{27}$$

Again, we isolate the power term ρ_{M-2} and obtain

$$\rho_{M-2} = \sqrt{\rho_{M-1}\phi + \frac{\rho_M \phi^2}{\rho_{M-1}} + \frac{M\lambda\phi^3}{\rho_{M-1}\rho_M}}. \tag{28}$$

We can continue this procedure for all $m \in \{1, \ldots, M\}$ and the results can be summarized as:

$$\rho_M = f(\lambda), \tag{29}$$
$$\rho_{M-1} = f(\lambda, \rho_M), \tag{30}$$
$$\rho_{M-2} = f(\lambda, \rho_M, \rho_{M-1}), \tag{31}$$
$$\vdots$$
$$\rho_1 = f(\lambda, \rho_M, \ldots, \rho_3, \rho_2). \tag{32}$$

By utilizing a method that is similar to the backward substitution approach [48 App. C.2], we can obtain a relationship between the power terms ρ_m as follows: first, by substituting $M\lambda\phi = \rho_M^2$ (see (24)) in (26) (or equivalently in (30)) we evaluate ρ_{M-1} as $\rho_{M-1} = \sqrt{2\phi\rho_M}$. Next, ρ_{M-2} is evaluated by substituting $\sqrt{M\lambda\phi} = \rho_M$ and $\sqrt{2\phi\rho_M} = \rho_{M-1}$ in (31). By continuing this procedure we can express the optimal transmit power in the m^{th} round as

$$\rho_m = \sqrt{2\phi\rho_{m+1}},\ 1 \leqslant m < M. \tag{33}$$

Since ρ_M is a function of λ (see (24)) and using (33), it is clear that each ρ_m is a function of λ. Thus, all that remains is to compute the Lagrangian multiplier λ. For this purpose, we utilize the outage constraint in (20) (**C4**). First, we substitute ρ_m for $m = 1, \ldots, M$ in (19) to obtain E_M^{ARQ} as

$$E_m^{ARQ} = \frac{\phi^M}{\prod_{m=1}^{M} \rho_m} = \epsilon, \tag{34}$$

where ρ_m can be found as

$$\rho_m = \sqrt{2\phi\sqrt{2\phi\sqrt{2\phi \ldots \sqrt{M\lambda\phi}}}} \tag{35}$$

We observe that in (35) we have product of nested roots. To obtain a simpler expression for ρ_m we can group the repetitive elements under a certain exponent. This would then result in

$$\rho_m = \sqrt{2^{a(m)}\phi^{b(m)}(M\lambda)^{c(m)}}. \tag{36}$$

In (36), we can compute the exponents

$$a(m) = 2 - 2^{-(M-m-1)}, \tag{37}$$
$$b(m) = 2 - 2^{-(M-m)}, \tag{38}$$
$$c(m) = 2^{-(M-m)}, \tag{39}$$

Finally, we compute λ by equating E_M^{ARQ} to the outage target ϵ based on **C4**:

$$\lambda = \left(\frac{\phi^{2M}}{\epsilon^2 \prod_{m=1}^{M} (2^{a(m)} \phi^{b(m)} M^{c(m)})} \right)^{l(m)}. \tag{40}$$

where $l(m) = \sum_{m=1}^{M} \frac{1}{c(m)}$.

In conclusion, for Type-I ARQ protocol the power allocation is done as shown in algorithm 1.

Algorithm 1 Power allocation for Type-I ARQ

1: **Inputs:** ϕ, M.
2: **Compute:** $a(m)$ as in (37).
3: **Compute:** $b(m)$ as in (38).
4: **Compute:** $c(m)$ as in (39).
5: **Compute:** λ as in (40).
6: **Compute:** ρ_M as in (24).
7: **while** all power terms are not found **do**
8: **Compute:** ρ_m as in (33).
9: **Decrease:** m by 1.
10: **end while**.
11: **Outputs:** All the power terms ρ_m.

3.1.2. CC-HARQ

Another member of the repetitive retransmission schemes is CC-HARQ protocol. This protocol is widely implemented in several standards (e.g. HSDPA, LAN etc). Its principle is illustrated in Figure 3.2. When utilizing this protocol, each transmission is done with full rate. Thus, the whole packet is sent in each round. However, unlike the previous protocol we analyzed, all the received packets are buffered and the receiver performs MRC to enhance the SNR. This results in combining gain and higher order diversity gain.

For this protocol, in [37] the authors have derived a very tight approximation for E_M as

$$E_M^{CC-HARQ} = \frac{\phi^M}{M! \prod_{m=1}^{M} \rho_m}. \tag{41}$$

To obtain the optimal power allocation scheme, we solve problem (20). We start our solution by writing the Lagrangian function, which is

$$\mathcal{L}(\rho_m, \mu_m, \lambda) = \frac{1}{M} \sum_{m=1}^{M} \rho_m E_{m-1}^{CC-HARQ} + \sum_{m=1}^{M} \mu_m \rho_m + \lambda (E_M^{CC-HARQ} - \epsilon), \tag{42}$$

where μ_m for $m = 1, \ldots, M$ and λ are the new Lagrangian multipliers. Next, we write the KKT conditions as

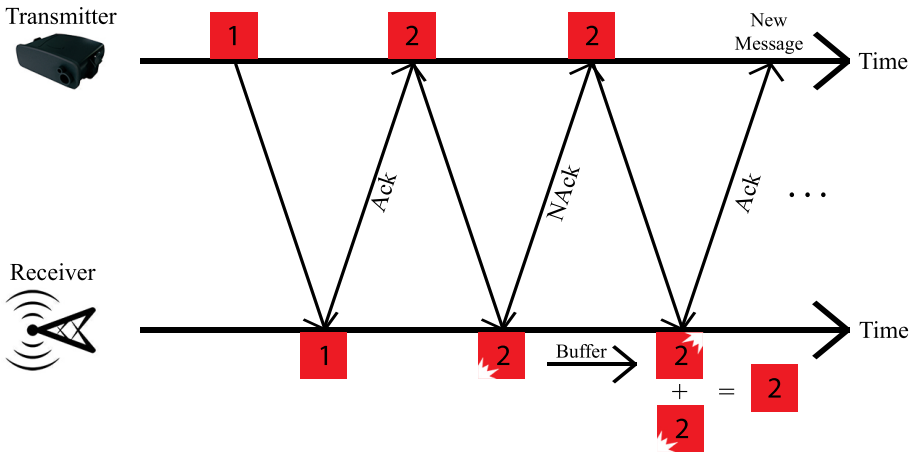


Figure 3.2: The setup for CC-HARQ protocol. The transmitter sends first packet 1 and waits for the Ack from the receiver. Then it sends packet 2. If the receiver can not decode the packet, it buffers the packet and asks for retransmission. Then, when it receives packet 2 once more, it combines it with the buffered packet to extract the information.

C1 $\frac{\partial \mathcal{L}}{\partial \rho_m} = 0,\ m = 1, \ldots, M,$

C2 $\mu_m \geqslant 0,\ m = 1, \ldots, M,$

C3 $\mu_m \rho_m = 0,\ m = 1, \ldots, M,$

C4 $E_M^{CC-HARQ} - \epsilon = 0.$

We can write the derivative of the Lagrangian function $\mathcal{L}(\rho_m, \mu_m, \lambda)$ with respect to the power ρ_m as

$$\frac{\partial \mathcal{L}(\rho_m, \mu_m, \lambda)}{\partial \rho_m} = \frac{1}{M}\left(\frac{\phi^{m-1}}{(m-1)!\prod_{i=1}^{m-1}\rho_i} - \sum_{i=1}^{M-m}\frac{\rho_{m+i}\phi^{m+i-1}}{(m+i-1)!\rho_m^2\prod_{j=1,j\neq m}^{m+i-1}\rho_j}\right) - \mu_m - \lambda\left(\frac{\phi^M}{M!\rho_m^2\prod_{i=1,i\neq m}^{M}\rho_i}\right). \tag{43}$$

Similarily to the ARQ protocol, we can argue that the inequality constraint is inactive. Therefore, using (43) and $\mu_M = 0$, we can write **C3** for $m = M$ as

$$\frac{\partial \mathcal{L}(\rho_m, \mu_m, \lambda)}{\partial \rho_M} = \frac{1}{M!}\frac{\phi^{M-1}}{\prod_{i=1}^{M-1}\rho_i} - \lambda\left(\frac{\phi^M}{\rho_M^2 M!\prod_{i=1,i\neq m}^{M}\rho_m}\right) = 0. \tag{44}$$

After some mathematical manipulations of (44) we obtain the transmit power at the M^{th} CC-ARQ round as

$$\rho_M = \sqrt{\frac{M\lambda\phi}{M}}. \tag{45}$$

Next, we substitute $m = M-1$ in (43) and utilize the fact that $\mu_{M-1} = 0$. Based on these, we can rewrite **C1** for $m = M-1$ as

$$\frac{\partial \mathcal{L}(\rho_m, \mu_m, \lambda)}{\partial \rho_{M-1}} = \frac{1}{M}\left(\frac{\phi^{M-2}}{(M-2)!\prod_{i=1}^{M-2}\rho_i} - \frac{\rho_M \phi^{M-1}}{(M-1)!\rho_{M-1}^2 \prod_{i=1, i\neq M-1}^{M-2}\rho_i}\right) - \lambda\left(\frac{\phi^M}{M!\rho_{M-1}^2\prod_{i=1, i\neq M-1}^{M}\rho_i}\right) = 0. \tag{46}$$

Furthermore, by making some algebraic simplifications in (46) we obtain the following expression for ρ_{M-1}

$$\rho_{M-1} = \sqrt{\frac{\rho_M \phi}{M-1} + \frac{\lambda\phi^2}{\rho_M(M-1)}}. \tag{47}$$

Again, we observe that we have obtained a similar structure to the one presented in the case of the ARQ protocol. This can be achieved by making some further manipulations to equations (45) and (47)

$$\rho_M = \sqrt{\phi\lambda}, \tag{48}$$

$$\rho_{M-1} = \sqrt{\frac{2\rho_M\phi}{M-1}}, \tag{49}$$

$$\rho_{M-2} = \sqrt{\frac{2\rho_{M-1}\phi}{M-2}}, \tag{50}$$

$$\vdots$$

$$\rho_1 = \sqrt{2\phi\rho_2}. \tag{51}$$

From the structure above, we observe that we can still apply the back-substitution approach and obtain ρ_m as

$$\rho_m = \sqrt{\frac{2\phi\rho_{m+1}}{m}}. \tag{52}$$

Now, all the power terms are computed as a function of λ. To find the value of the equality Lagrange multiplier we utilize **C4** and write:

$$E_M^{CC-HARQ} = \frac{\phi^M}{M!\prod_{m=1}^{M}\rho_m} = \epsilon, \tag{53}$$

where ρ_m is calculated as

$$\rho_m = \frac{2^a\phi^b\lambda^c}{\sqrt{m}}. \tag{54}$$

In (54), the exponents $a(m)$, $b(m)$ and $c(m)$ are defined in (37), (38) and (39). Finally, we compute the Lagrangian multiplier λ as:

$$\lambda = \left(\frac{\phi^{2M}}{\epsilon^2}\prod_m \frac{1}{2^{a(m)}\phi^{b(m)}}\right)^{l(m)}. \tag{55}$$

where $l(m) = \sum_{m=1}^{M}\frac{1}{c(m)}$.

In conclusion, for CC-HARQ protocol the power allocation is done as shown in algorithm 2.

Algorithm 2 Power allocation for CC-HARQ

1: **Inputs:** ϕ, M.
2: **Compute:** $a(m)$ as in (37).
3: **Compute:** $b(m)$ as in (38).
4: **Compute:** $c(m)$ as in (39).
5: **Compute:** λ as in (55).
6: **Compute:** ρ_M as in (48).
7: **while** all power terms are not found **do**
8: **Compute:** ρ_m as in (52).
9: **Decrease:** m by 1.
10: **end while**.
11: **Outputs:** All the power terms ρ_m.

3.1.3. Example for M=2 and a simplified solution

In this subsection present a simplified method for obtaining the optimal power terms, which is valid for the case when $M = 2$. Here, we make detailed derivations only for the Type-I ARQ protocol. Similarily, the expressions for the power terms can be derived for the case of CC-HARQ protocol.

For the scenario of $M = 2$ transmissions the optimization problem (20) simplifies to

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}(\rho_1 + \rho_2\epsilon_1) \\ &\text{subject to} \quad \frac{\phi^2}{2\rho_1\rho_2} = \epsilon \end{aligned} \tag{56}$$

To solve problem (56) we can utilize Algorithm. 1, derived in Section 3.1.1. First, we compute the exponents $a(m), b(m)$ and $c(m)$ as shown in (37), (38) and (39). Next, we find the value of λ from (40) as

$$\lambda = \frac{\phi}{2\epsilon}\sqrt[3]{\frac{1}{2\epsilon}}. \tag{57}$$

Further, by using (24) and (33) we compute the power terms as:

$$\rho_2 = \sqrt{2\frac{\phi}{2\epsilon}\sqrt[3]{\frac{1}{2\epsilon}}\,(e^R - 1)}, \tag{58}$$

$$\rho_1 = \sqrt{2\,(e^R - 1)\sqrt{2\frac{\phi}{2\epsilon}\sqrt[3]{\frac{1}{2\epsilon}}\,(e^R - 1)}}, \tag{59}$$

Mathematical simplification of (58) and (59) would lead to the final values of the power terms as

$$\rho_1 = \phi\sqrt[3]{\frac{1}{\epsilon}}, \tag{60}$$

$$\rho_2 = \frac{\phi}{2\epsilon}\sqrt[3]{\epsilon}. \tag{61}$$

For this specific case of $M = 2$ we can also utilize the following simpler approach to find the optimal power allocation. First, we rewrite the equality constraint as $\rho_2 = \frac{\phi^2}{2\rho_1\epsilon}$. Next, by substituting ρ_2 in the objective function of (56), we obtain an unconstrained optimization problem with variable ρ_1 as

$$\text{minimize} \quad \frac{1}{2}\left(\rho_1 + \frac{\phi^3}{\rho_1^2\epsilon}\right). \tag{62}$$

Then, we compute ρ_1 by setting the first derivative of the new objective function to zero. This results in

$$1 - \frac{2\phi^3}{\rho_1^3\epsilon} = 0. \tag{63}$$

Finally, by solving (89) for ρ_1 we find $\rho_1 = \phi\sqrt[3]{\frac{2}{\epsilon}}$, which is same as what we obtained by using the procedure described in Section 3.1.1. Then, after substituting the first power term equation in the rewritten equality constraint we compute $\rho_2 = \frac{\phi}{\epsilon}\sqrt[3]{\frac{\epsilon}{2}}$.

After applying the same procedure for the case of CC-HARQ protocol, the power terms would be

$$\rho_1 = \frac{\phi}{\sqrt[3]{\epsilon}}, \tag{64}$$

$$\rho_2 = \frac{\phi}{2\epsilon}\sqrt[3]{\epsilon}. \tag{65}$$

3.2. Parallel channel coding schemes

Here, we present the power allocation algorithm for parallel channel coding schemes. In this case, the transmitter sends different and jointly designed packets for each of the messages. There are several protocols which implement this scheme, but in the thesis we focus on Incremental Redundancy (IR) HARQ protocols. The reason for that is because recently an advanced Low Density Parity Check (LDPC) channel coding scheme has been adopted for 5G new radio air interface. It provides full rate compatibility with IR-HARQ, flexibility for blocklength scaling and finite rate granularity [49].

In the family of IR-HARQ the most popular protocols are Fixed Rate (FR) IR-HARQ and Variable Rate (VR) IR-HARQ. In the first case, all the packets that are transmitted have the same number of nats in them. On the other hand, when utilizing VR IR-HARQ the size of the transmitted packets can differ. It can be an optimization parameter, and can adapt to the channel conditions, transmitted power, channel coding rate that might be needed for a certain round.

Herein, we analyze FR IR-HARQ, which is also the final retransmission protocol that we analyze in the thesis. As shown in Figure 3.3, when implementing this protocol the transmitter splits the parent codeword into M sub-codewords of length τ. Each of these codewords is transmitted during one round. In this case, the transmission rate would be $R = \frac{K}{\tau}$. This will result in an improve of spectral efficiency, as will be shown in Section 4.1. The receiver buffers all the received packets and utilizes them to recover the information.

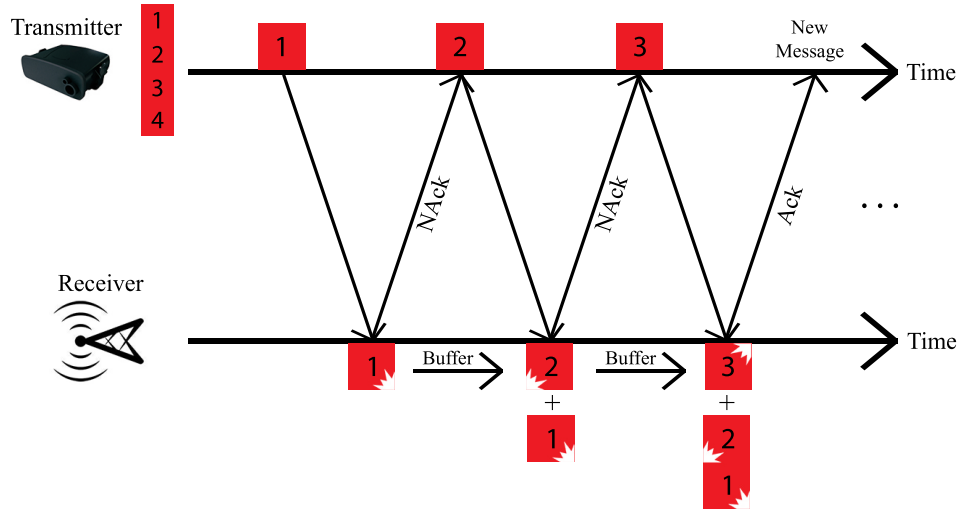


Figure 3.3: The setup for IR-HARQ protocol. The transmitter splits the message into several packets. It sends first packet 1 and waits for reply from the receiver. If it receives NAck packet, then it sends more redundancy through packet 2 and so on. Meanwhile, the receiver buffers the packets and combines them to extract the information.

Let us revisit the results for long codewords using a Gaussian codebook. In this case, the mutual information that is gathered would be

$$I = \frac{1}{M} \sum_{m=1}^{M} \log\left(1 + \rho_m\right). \tag{66}$$

After making some manipulations we can compute the outage probability as

$$E_M^{IR-HARQ} = \Pr\left[\sum_{m=1}^{M} \log\left(1 + \rho_m\right) < MR\right]. \tag{67}$$

In [50] the author provides a theorem which can be used to approximate (67). Therein an integral form approximation of the outage probability is provided. Herein, we resort to those results and provide closed-form approximation for the outage probability when different power levels are allocated in each IR-HARQ round. It can be computed as

$$E_M^{IR-HARQ} = \frac{g_M}{\prod_{m=1}^{M} \rho_m}. \tag{68}$$

The proof is shown in Appendix. 7.3. Again, to compute the power allocation algorithm, we solve problem (20). First, we write the Lagrangian function as

$$\mathcal{L}(\rho_m, \mu_m, \lambda) = \frac{1}{M} \sum_{m=1}^{M} \rho_m E_{m-1}^{IR-HARQ} + \sum_{m=1}^{M} \mu_m \rho_m + \lambda(E_M^{IR-HARQ} - \epsilon), \tag{69}$$

where μ_m for $m = 1, \ldots, M$ and λ are the new Lagrangian multipliers. Next, we write the KKT conditions as

C1 $\frac{\partial \mathcal{L}}{\partial \rho_m} = 0,\ m = 1, \ldots, M,$

C2 $\mu_m \geqslant 0,\ m = 1, \ldots, M,$

C3 $\mu_m \rho_m = 0,\ m = 1, \ldots, M,$

C4 $E_M^{IR-HARQ} - \epsilon = 0.$

We can write the derivative of the Lagrangian function $\mathcal{L}(\rho_m, \mu_m, \lambda)$ with respect to the power ρ_m as

$$\frac{\partial \mathcal{L}(\rho_m, \mu_m, \lambda)}{\partial \rho_m} = \frac{1}{M}\left(\frac{g_{m-1}}{\prod_{i=1}^{m-1}\rho_i} - \sum_{m=1}^{M-m}\frac{\rho_{m+i}g_{m+i-1}}{\rho_m^2\prod_{i=1,i\neq m}^{m+i-1}\rho_i}\right) - \mu_m - \lambda\left(\frac{g_M}{\rho_m^2\prod_{i=1,i\neq m}^{M}\rho_i}\right). \tag{70}$$

Again, we argue that the inequality constraint is inactive. Furthermore, following a similar approach as in the protocols which utilize repetition coding, we write the power term for the last round as

$$\frac{\partial \mathcal{L}(\rho_m, \mu_m, \lambda)}{\partial \rho_M} = \frac{1}{M}\frac{g_{M-1}}{\prod_{i=1}^{M-1}\rho_i} - \lambda\left(\frac{g_M}{\rho_M^2\prod_{i=1,i\neq M}^{M}\rho_m}\right) = 0. \tag{71}$$

Next, we make some algebraic gymnastics of (71) and obtain the transmit power at the M^{th} round as

$$\rho_M = \sqrt{\frac{M\lambda g_M}{g_{M-1}}}. \tag{72}$$

Then, we compute the power term before the last. To do this, we substitute $m = M - 1$ in (70) and rewrite **C1** as

$$\frac{\partial \mathcal{L}(\rho_m, \mu_m, \lambda)}{\partial \rho_{M-1}} = \frac{1}{M}\left(\frac{g_{M-2}}{\prod_{i=1}^{M-2}\rho_i} - \frac{\rho_M g_{M-1}}{\rho_{M-1}^2\prod_{i=1,i\neq M-1}^{M-2}\rho_i}\right) - \lambda\left(\frac{g_M}{\rho_{M-1}^2\prod_{i=1,i\neq M-1}^{M}\rho_i}\right) = 0. \tag{73}$$

Furthermore, by making some simplifications of (73) we obtain the expression for the power term ρ_{M-1} as

$$\rho_{M-1} = \sqrt{\frac{1}{g_{M-2}}\left(\rho_M g_{M-1} + \frac{M\lambda g_M}{\rho_m g_{M-1}}\right)} \tag{74}$$

By continuing this process and making further mathematical manipulations, we notice that same as in the repetitive schemes, we can obtain the following structure for the parallel retransmission schemes.

$$\rho_M = \sqrt{\frac{\lambda M g_M}{g_{M-1}}}, \tag{75}$$

$$\rho_{M-1} = \sqrt{\frac{2\rho_M g_{M-1}}{g_{M-2}}}, \tag{76}$$

$$\rho_{M-2} = \sqrt{\frac{2\rho_{M-1} g_{M-2}}{g_{M-3}}}, \tag{77}$$

$$\vdots$$

$$\rho_1 = \sqrt{\frac{2\rho_2 g_1}{g_0}}. \tag{78}$$

Again, we notice that the back-substitution approach is applicable, and the power term of the m^{th} round can be written as

$$\rho_m = \sqrt{\frac{2\rho_{m+1} g_m}{g_{m-1}}}. \tag{79}$$

At this point, we can compute each of the power terms as a function of the equality Lagrangian multiplier, λ. To derive a closed-form expression for it, we utilize **C4**

$$E_M^{IR-HARQ} = \frac{g_M}{M! \prod_{m=1}^{M} \rho_m} = \epsilon, \tag{80}$$

where the power term in the m^{th} round can be found as

$$\rho_m = 2^{a(m)} (m\lambda)^{c(m)} \prod_{i=1}^{m} \frac{g_i}{g_{i-1}}^{d(i)}. \tag{81}$$

In (81) $a(m)$ and $c(m)$ can be found from (37) and (38), respectively. Further, the exponent $d(i)$ is found as

$$d(i) = 2^{-i}. \tag{82}$$

Finally, we can find the value of λ as

$$\lambda = \left(\frac{g_M}{\epsilon^2} \prod_{m=1}^{M} \frac{1}{2^{a(m)} M^{c(m)} \prod_{i=1}^{m} \left(\frac{g_i}{g_{i-1}} \right)^{d(i)}} \right)^{l(m)}. \tag{83}$$

where $l(m) = \sum_{m=1}^{M} \frac{1}{c(m)}$.

For this protocol, the power allocation algorithm is given below

Algorithm 3 Power allocation for IR-HARQ

1: **Inputs:** R, M.
2: **while** all values of $g_m(R)$ are not found **do**
3: **Compute:** g_m as in (136).
4: **Increase:** m by 1.
5: **end while**.
6: **Compute:** $a(m)$ as in (37).
7: **Compute:** $c(m)$ as in (39).
8: **Compute:** $d(i)$ as in (82).
9: **Compute:** λ as in (83).
10: **Compute:** ρ_M as in (75).
11: **while** all power terms are not found **do**
12: **Compute:** ρ_m as in (81).
13: **Decrease:** m by 1.
14: **end while**.
15: **Outputs:** All the power terms ρ_m.

3.2.1. Example for M=2 and a simplified solution

In this subsection present a simpler method for obtaining the optimal power terms in the case when the maximum number of transmissions is set to two. Same as in Section 3.1.3 optimization problem (20) simplifies to

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}(\rho_1 + \rho_2\epsilon_1) \\ &\text{subject to} \quad \frac{g_2(R)}{\rho_1\rho_2} = \epsilon \end{aligned} \tag{84}$$

To solve problem (56) we can utilize Algorithm 3, derived in Section 3.2. First, we utilize the term $g_2(R)$, which is given in (132). Next, we compute exponents $a(m), c(m)$ and $d(m)$ as shown in (37), (39) and (82). Further, from (83) we find the equality Lagrange multiplier as

$$\lambda = \left(\frac{g_2}{\epsilon^2} \prod_{m=1}^{2} \frac{1}{2^{a(m)+c(m)} \prod_{i=1}^{m} \frac{g_i}{g_{i-1}}^{d(i)}} \right)^{l(2)}. \tag{85}$$

Further, by using (75) and (81) and making some algebraic simplifications we compute the power terms as:

$$\rho_1 = \sqrt[3]{\frac{2(1 - e^R + Re^R)(e^R - 1)}{\epsilon}}, \tag{86}$$

$$\rho_2 = \sqrt[3]{\frac{(1 - e^R + Re^R)^2}{2\epsilon^2(e^R - 1)}}. \tag{87}$$

For this specific case of $M = 2$ we can also utilize the following simpler approach to find the optimal power allocation. First, we rewrite the equality constraint

as $\rho_2 = \frac{g_2(R)}{\rho_1 \epsilon}$. Next, by substituting ρ_2 in the objective function of (84), we obtain an unconstrained optimization problem with variable ρ_1 as

$$\text{minimize} \quad \frac{1}{2}\left(\rho_1 + \frac{\mathrm{g_1 g_2}}{\rho_1^2 \epsilon}\right). \tag{88}$$

Then, we compute ρ_1 by setting the first derivative of the new objective function to zero. This results in

$$1 - \frac{2(1 - e^R + Re^R)(e^R - 1)}{\rho_1^3 \epsilon} = 0. \tag{89}$$

Finally, by solving (89) for ρ_1 we find $\rho_1 = \sqrt[3]{\frac{2(1-e^R+Re^R)(e^R-1)}{\epsilon}}$, which is same as what we obtained by using the procedure described in Section 3.2. Then, after substituting the first power term equation in the rewritten equality constraint we compute $\rho_2 = \sqrt[3]{\frac{(1-e^R+Re^R)^2}{2\epsilon^2(e^R-1)}}$, which as we can see from (83) and (84) match with the proposed solution.

4. THROUGHPUT AND DELAY ANALYSIS

Utilizing schemes that are based on retransmissions causes the latency of a system to increase due to the presence of feedback and information repetition. Furthermore, these schemes are also associated with a degradation of the spectral efficiency. These effects, are then reflected in the overall system throughput. In this chapter we derive the closed form expressions for the throughput of the retransmission protocols presented above for the general case of M transmissions. Furthermore, we evaluate mathematically the delay that is introduced in the system by these protocols.

4.1. Throughput analysis

In the open loop setup, which was analyzed in Section 2.2, the throughput (η) can be computed as

$$\eta = R\left(1 - \epsilon(n, R, \rho)\right). \tag{90}$$

However, during our analysis we need to consider the fact that after each retransmission round there exists a degradation of spectral efficiency. As a result of this degradation the spectral efficiency up to the M^{th} retransmission round is given by

$$\omega_M = \frac{R}{M} = \frac{K}{Mn}. \tag{91}$$

From (91) we observe the throughput gain that parallel schemes have over repetition schemes. For easier illustration of this statement, lets assume that we have one on one mapping between the number of information nats and the number of channel uses. Furthermore, lets fix the maximum number of transmissions to $M = 3$. As mentioned above, in the case of parallel retransmission schemes, the parent codeword is split into smaller sub-codewords, which are then transmitted in each round. On the other hand, when repetition schemes are utilized, the entire packet is transmitted. If all three rounds are exhausted, the spectral efficiencies for the parallel and repetition schemes would be $\omega_3^{par} = 1$ ncpu and $\omega_3^{rep} = \frac{1}{3}$ ncpu, respectively.

This observation can be better illustrated using Figure. 4.1. To obtain the the plot, we assume that we need to convey 500 information nats, which are mapped in a packet of size 500 channel uses. We set $M = 5$. Therefore, when parallel schemes are utilized, the information will be conveyed in 5 packets of size 100 channel uses. On the other hand, to convey the information with repetition schemes, 5 transmissions of packets of size 500 channel uses are needed. From the figure, we can notice that the open loop setup serves as asymptotic lower bound for the parallel scheme, and as upper limit for the repetition scheme.

To obtain the closed form expression for the throughput, we first need to compute the total number of channel uses. For the case of the parallel schemes, the codeword of length n, will be divided into M sub-codewords of length τ in a way that $n = M\tau$. On the other hand, for the case of the repetition schemes, the total number of channel uses, ξ, until the last round would be $\xi^{rep} = Mn$. If the the packet is successfully decoded

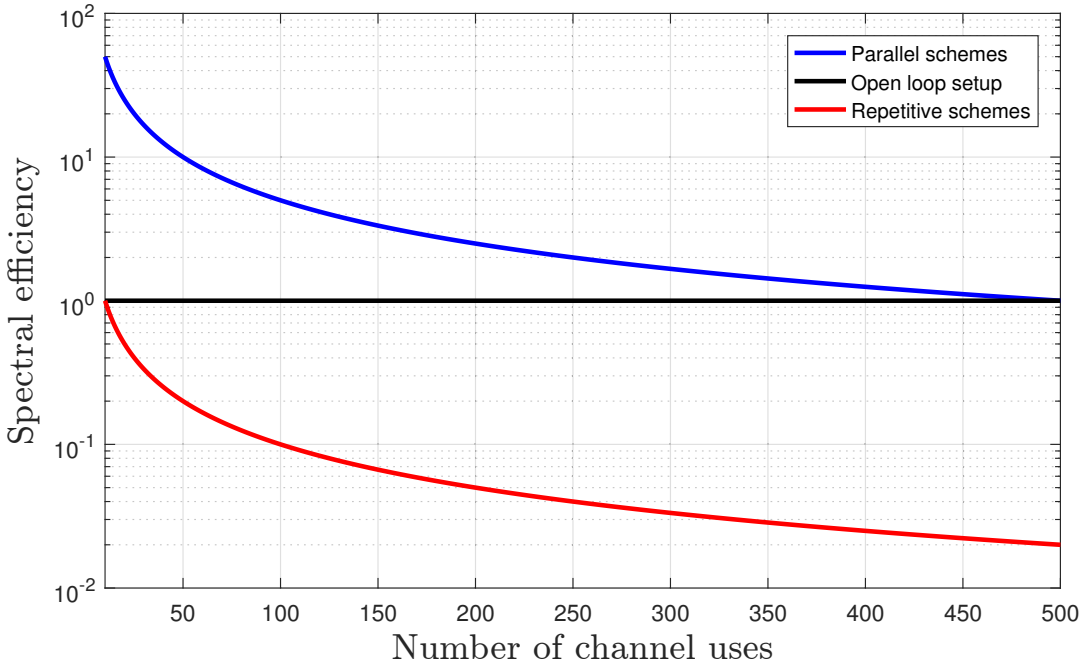


Figure 4.1: Spectral efficiency of retransmission schemes as a function of the number of channel uses.

in the m^{th} round, then we can find the total number of channel uses for the parallel schemes as

$$\begin{cases} \xi^{par} = m\left(\tau + D\right), & if \quad 0 \leqslant m < M \\ \xi^{par} = n + MD, & if \quad m = M \end{cases} \tag{92}$$

where D denotes the delay of the feedback transmission expressed in channel uses. Note that D accounts for the feedback delay from the NACK transmission, which are also assumed to use a few hundred channel uses. As pointed out in [4], this models more accurately the impact of the retransmission protocols in practical scenarios compared to the conventional one nat feedback. In the case of the repetition retransmission schemes the total number of channel uses would be

$$\begin{cases} \xi^{rep} = m\left(n + D\right), & if \quad 0 \leqslant m < M \\ \xi^{rep} = M\left(n + D\right). & if \quad m = M \end{cases} \tag{93}$$

These equations are based on the fact that in each retransmission round an acknowledgment (ACK) or non acknowledgment (NACK) packet is sent back from the receiver. In the M^{th} round, since the transmission will be interrupted, there will be no feedback from the receiver. To compute the expected number of channel uses in each packet transmission, we must bear in mind that only the first transmission is sure to happen. The retransmissions will happen only if the previous transmission has been unsuccessful. Based on this, we can find the expected number of channel uses in each packet transmission as

$$\mathcal{T} = \sum_{m=1}^{M} mnE_{m-1} + D\sum_{m=1}^{M-1} mE_{m-1}, \tag{94}$$

where E_m is the packet drop probability given in (19) and $E_0 = 1$.

If the packet is successfully decoded in any round, all K information nats are recovered by the receiver. Therefore, we can compute the expected number of information nats that will be transmitted as

$$\mathcal{K} = K(1 - E_M) = K\left(1 - \prod_{m=1}^{M} \epsilon_m\right). \tag{95}$$

Finally, following the renewal-reward theorem [51, 52], the throughput for the case of M transmissions can be expressed as:

$$\eta = \frac{\mathcal{K}}{\mathcal{T}} = \frac{K\left(1 - \prod_{m=1}^{M} \epsilon_m\right)}{\sum_{m=1}^{M} mnE_{m-1} + D\sum_{m=1}^{M-1} mE_{m-1}}. \tag{96}$$

It is obvious from (96) that the presence of feedback causes loss in the throughput. This degradation becomes worse as we increase the number of retransmissions. However, the proposed power allocation scheme helps to mitigate this effect as will be shown in the next chapter.

Another important observation that we can make from the equation above is the intuition behind parallel retransmission schemes. If the channel conditions are not good, then all the transmissions are exhausted. This would give a performance which is very close to that of the open loop setup (slightly worst as the feedback delays will have to be accounted as well). However, if the channel conditions are good, then the throughput gains would be quite large. The cost of this gambling is an increase in the overall system complexity.

On the other hand, the situation for the repetition schemes is totally different. In that case, the higher the number of transmissions is, the more severe the throughput degradation will become. Furthermore, since each transmission has to contain the entire packet, repetition schemes will require very large buffer size in the receiver. However, their complexity is in general much lower with respect to parallel coding schemes.

It is clear that when comparing them to the parallel schemes there is the classical trade-off between the system throughput and complexity. For example, in applications in which throughput is not of vast importance, it would be more feasible to utilize repetition schemes. This is further reinforced by the fact that the nodes of such systems generally are very simple sensors. Implementing complex logic in them, would result in a large increase of their cost and would severely affect their battery lifetime, which in these applications corresponds to the device lifetime. However, in other applications the data volumes to be transmitted are far larger, and the throughput is important. Furthermore, the nodes communicating in these systems can afford extra complexity. Therefore, in these type of applications the utilization of parallel schemes would give a better performance.

4.2. Delay analysis

One of the most crucial elements in the systems that have ultra reliability requirements is the end-to-end delay. All the examples mentioned in the introduction are

delay-sensitive. However, the delay tolerance of the system highly depends on the application. Here, we try to evaluate the delay that is introduced by the utilization of the retransmission schemes.

Bounds on the outage probability and feedback delay for IR-HARQ schemes have been reported in [46]. There, the authors assume to have fixed and equal power allocation between the transmission rounds. For a certain power level, they find the feedback delay range that maximizes the throughput. Based on their assumptions, $D < n \cdot m \cdot r$, where

$$r = \frac{1 - \frac{1}{M}\sum_{m=1}^{M} E_{m-1}}{\sum_{m=1}^{M} E_{m-1}}. \tag{97}$$

Furthermore, by utilizing the bounds on the outage probability, the values of r are limited within the range

$$\frac{mn\left(M - 1 - \sum_{m=1}^{M-1} u_m\right)}{M\left(1 + \sum_{m=1}^{M-2} u_m\right)} \leqslant r \leqslant \frac{mn\left(M - 1 - \sum_{m=1}^{M-1} v_m\right)}{M\left(1 + \sum_{m=1}^{M-2} v_m\right)} \tag{98}$$

where the functions u_m and v_m can be found as

$$u_m = 1 - \frac{e^{-\kappa}}{+} e^{-\psi} 2 + \frac{1}{2} e^{\alpha_m} \rho^{-\epsilon m \tau} \Gamma\left(1 - \epsilon m\tau, \psi + \frac{1}{\rho}\right), \tag{99}$$

$$v_m = \frac{1}{2}\left(1 - \operatorname{erf}\left(\frac{-\kappa\delta}{\sqrt{2}}\right)\right) - e^{\frac{1-2\kappa\delta^2}{2\delta^2}}\left(1 - \operatorname{erf}\left(\frac{1-\delta^2\kappa}{\sqrt{2}\delta}\right)\right). \tag{100}$$

In the equations above we can find ψ and α as

$$\psi = \frac{e^{\frac{K}{m\tau} + \frac{\epsilon}{2}} - 1}{\rho}, \tag{101}$$

$$\alpha = \frac{1}{\rho} + K\epsilon + \frac{m\tau\epsilon^2}{2}. \tag{102}$$

As discussed in Section 4.1, when an URC setup is utilized, finding the range of delays that maximize the throughput is not always of vast importance. Generally, the parameter that should be optimized (e.g. outage probability, power, sum rate etc) is highly dependent on the application. However, in such setups, it is always important to have minimal delay. Therefore, it would be of high interest to analyze the behavior of the end-to-end delay of such systems depending on the maximum number of transmissions. We can obtain the mathematical expression for the delay from (96) as

$$D = \frac{\eta n \sum_{m=1}^{M} m E_{m-1} - K\left(1 - \prod_{m=1}^{M} \epsilon_m\right)}{\sum_{m=1}^{M} E_{m-1}} \tag{103}$$

Analytical minimization of (103) with respect to the number of transmissions would result result in very complex and intractable results. For this reason, we resort to numerical methods to find the delay-minimizing number of transmissions. The result is shown in Figure. 4.2. To obtain the plot, we fix both the throughput and channel coding

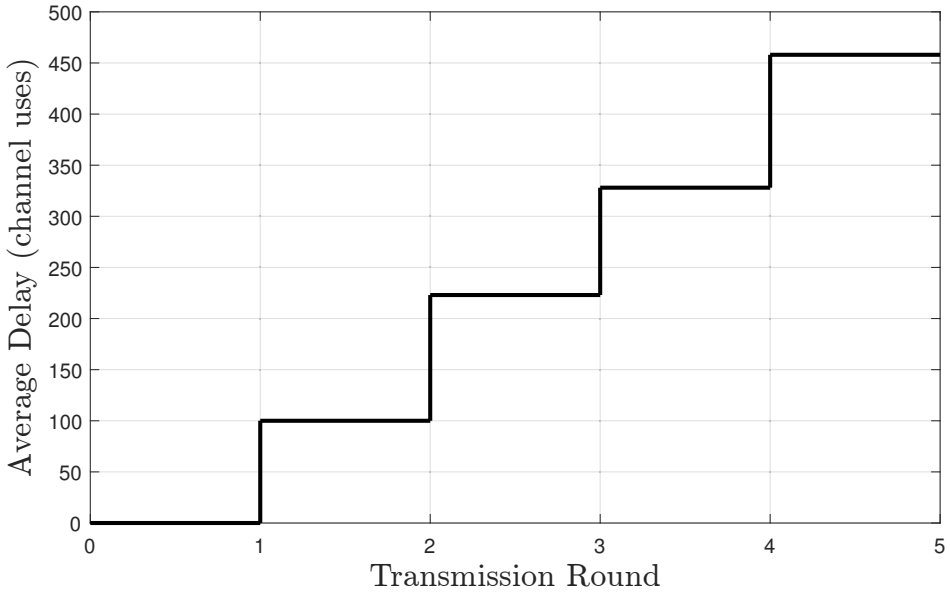


Figure 4.2: Average delay that will be introduced in the system after each transmission round when $\eta = R = 1$ ncpu.

rate to be 1 ncpu. Furthermore, for easier illustration, we assume to have deployed a power allocation scheme according to which the outage probability decreases by 30% with each transmission.

The results we obtain from the above plot allow us to quantify the delay that we will have in our system, based on the number of transmissions. Furthermore, it is obvious that there is no optimal delay-minimizing number of transmissions. What we notice is that the lower the number of transmissions, the lower the delay will be. This result fully matches the intuition about the problem. However, if the number of transmissions is lowered, the outage probability increases. Therefore, it is clear that there is a trade-off which will need further investigation.

5. NUMERICAL ANALYSIS

In this chapter we provide further results for the performance of the algorithms that we have developed during the thesis. First, we evaluate the power efficiency of the algorithms. We show how the power terms that will be transmitted in each round behave as a function of the outage probability. Further, we show that the utilization of the power allocation algorithms provides large gains when compared to the open loop setup. Also, we show that as the maximum number of transmission increases, the average power that is spent decreases. Next, we evaluate the impact of the power allocation algorithms in the throughput of the system. We show that the parallel schemes provide higher throughput when compared to the repetitive schemes. Finally, we show how the throughput is affected by the feedback delay.

5.1. Power efficiency

In Fig. 5.1 we illustrate the variation of transmit power ρ_m in each round versus the outage target ϵ for the case when we have a maximum of two transmissions. The channel coding rate is set to $R = 1$ ncpu. The results are obtained by implementing the equations derived in sections 3.1.3 and 3.2.1. The first observation we can make is that the IR-HARQ protocol gives the best performance in terms of saving power. Further, we notice that despite the protocol that is implemented, both power terms are lower than the open loop transmission which is shown in Fig.2.3. Moreover, if the first round is successful (i.e when the channel conditions are good), then the power gain with respect to the open loop setup would be very large. We earn around $20 - 25$ dB (depending on the protocol) for $\epsilon = 10^{-3}$, which corresponds to the start of the ultra reliable region. Then, the more stringent the reliability requirements are, the more power we save with respect to the open loop setup. We observe that in this region, the first power term is lower than the second power term. This holds even for the general case of M transmissions and the mathematical proof is shown in Appendix 7.1. Notice that this result fully matches the intuition. Since our goal is to achieve a target outage probability by spending as little power as possible and there is no delay limitation, we transmit first with low power. If the channel conditions are good then the transmission will be successful, and a large amount of power is saved. If it fails, then retransmissions are carried out until an Ack is received, or the maximum allowed number of retransmissions is reached.

Next, in Figure 5.2 we evaluate the average power as in (18) for each proposed scheme and for the scenario when the number of transmissions is increased. To attain this figure, we set $M = 3$ and $R = 1$. From it, we notice that the amount of power that is saved on average is significant, especially when compared to the open loop setup. Further, by comparing the result of Figure 5.2 with the results in Figure 5.1, we notice that as we increase the number of transmissions, we save more power on average.

In Figure 5.3 we evaluate the maximum power expenditure of our protocols in the case of $M \in \{2, 3\}$ transmissions and fixed channel coding rate $R = 1$ ncpu. To obtain the plot, we assume to have a "worst case scenario", where all the transmissions are exhausted. From it, we observe that the proposed algorithms again allow us to save power when compared to the open loop setup. The largest power gains, are again

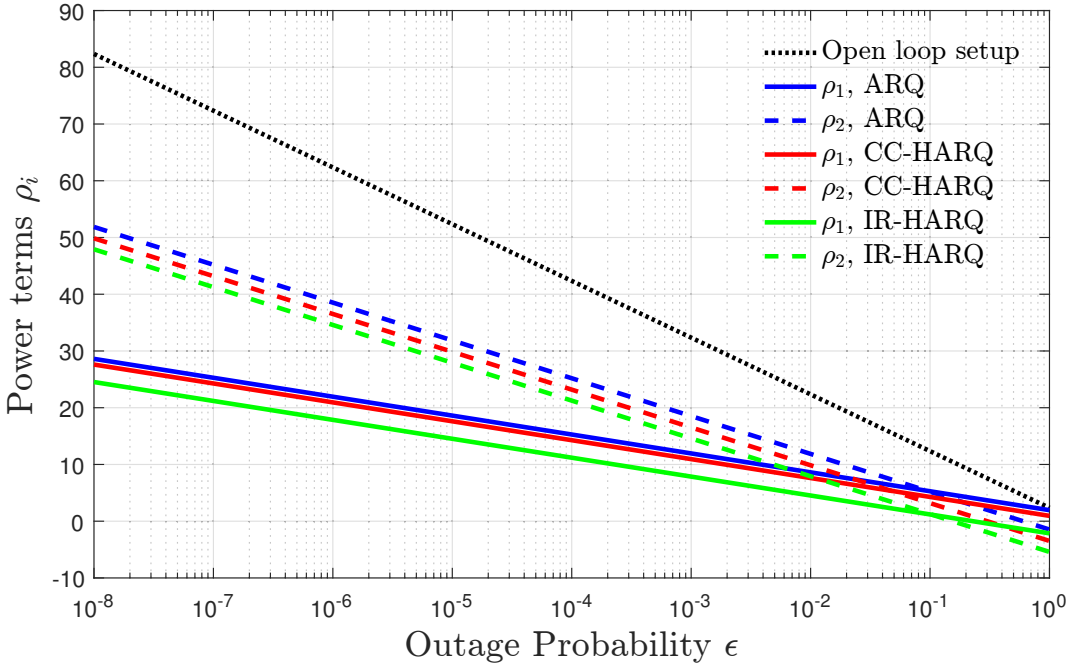


Figure 5.1: Transmit power in each round to achieve a target outage probability for rate $R = 1$ ncpu.

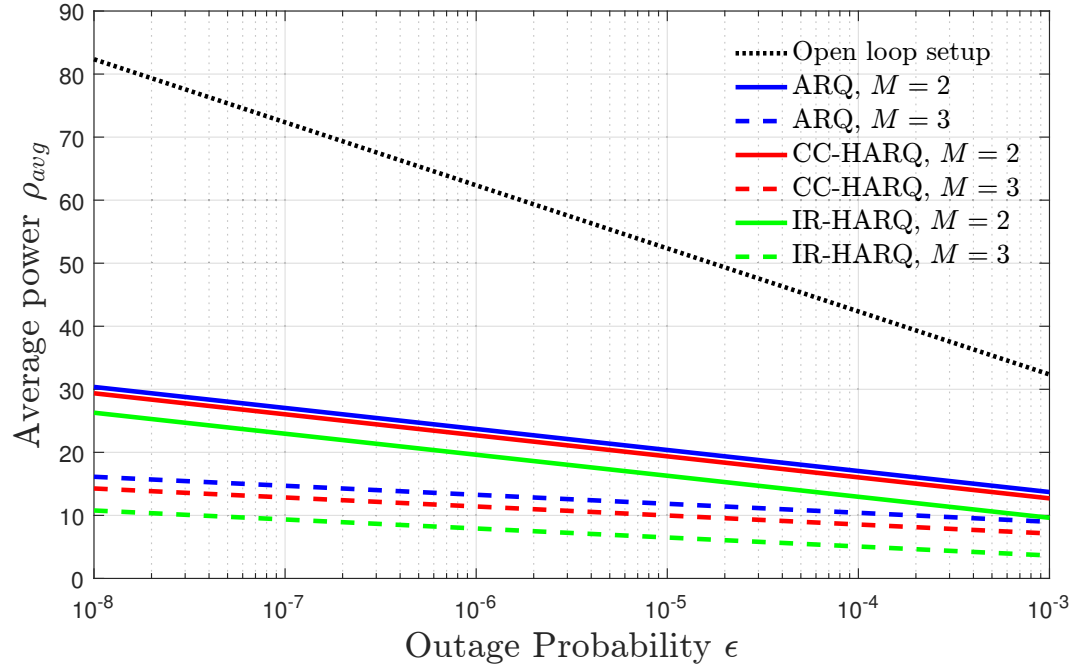


Figure 5.2: Average power required to achieve a target outage probability for rate $R = 1$ ncpu when the maximum number of transmissions is fixed to $M = 2$ and $M = 3$.

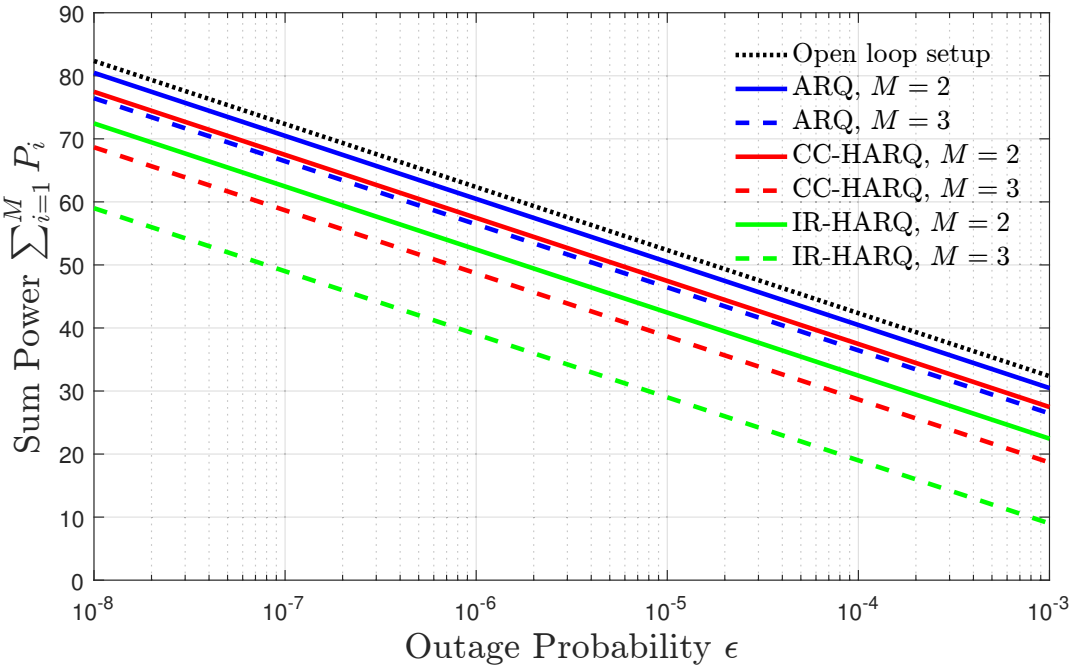
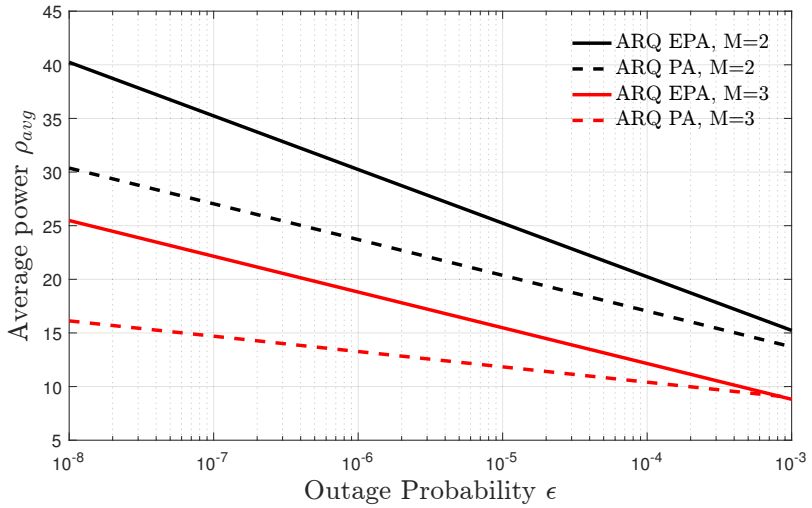


Figure 5.3: Maximum power that will be spent to achieve a target outage probability for rate $R = 1$ ncpu when the maximum number of transmissions is fixed to $M = 2$ and $M = 3$.

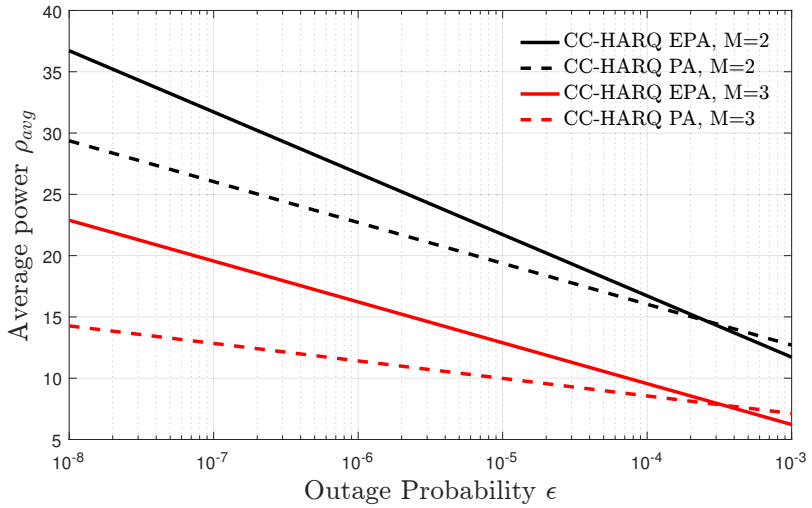
given from the IR-HARQ protocol. Notice that when $M = 3$, we can save over 20 dB by implementing this protocol. Furthermore, we observe that as we increase the number of transmissions, we will save power. This happens due to the fact that the diversity and combining gains become higher. Also, notice that in the case of Type-I ARQ protocol, the power gain when the number of transmissions increases is not as large as the other two protocols. This happens due to the fact that Type-I ARQ does not benefit from combining gains.

After showing the gains of our algorithms with respect to the open loop setup, we then compare them to a suboptimal power allocation scheme. We select the equal power allocation (EPA) because it provides a good benchmark and the comparison is shown in Figure 5.4. The channel coding rate is set to $R = 1$ ncpu and the plot is attained by setting $M = 2$ and $M = 3$. We notice that our proposed power allocation schemes allow us to save power in all possible scenarios. The gains with respect to EPA are around 10dB.

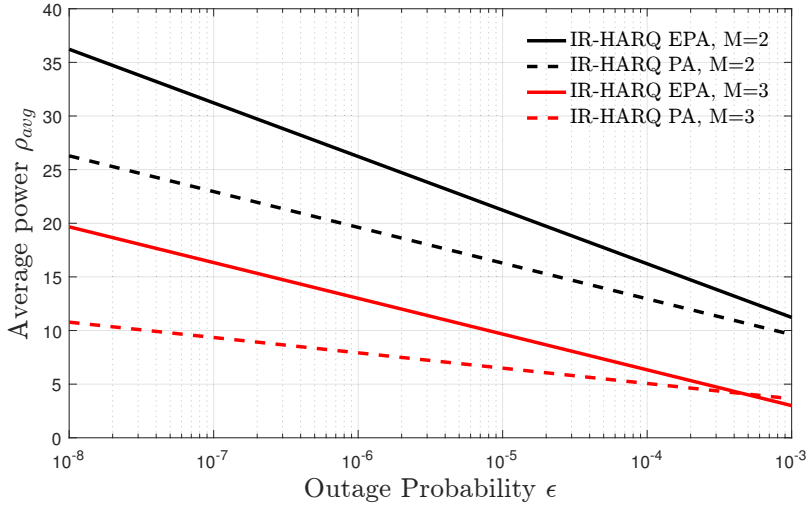
Finally, in Figure 5.5 we illustrate the behavior of the power terms as a function of the number of channel uses when $M = 2$ for each of the three protocols. Here we set the number of information nats K to 100 and 200. First, we notice that both power terms decrease as we increase the number of channel uses. Secondly we observe that when we increase the coding rate, we have to transmit with higher power in each round.



(a) Type-I ARQ

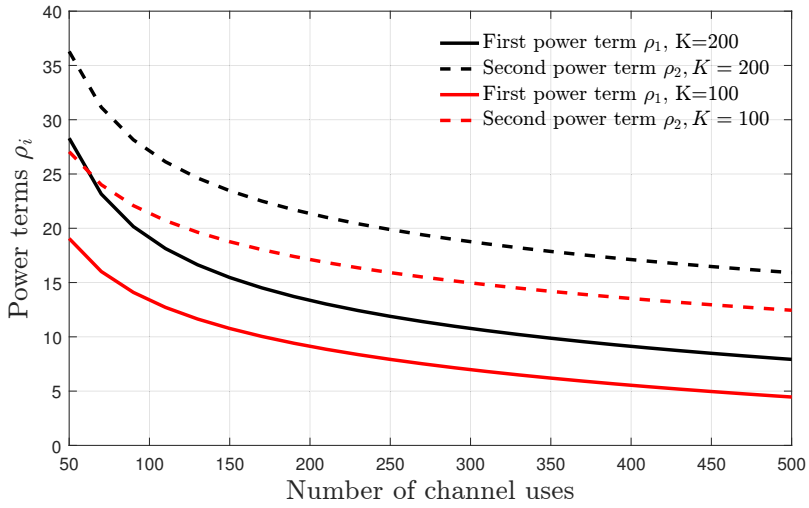


(b) CC-HARQ

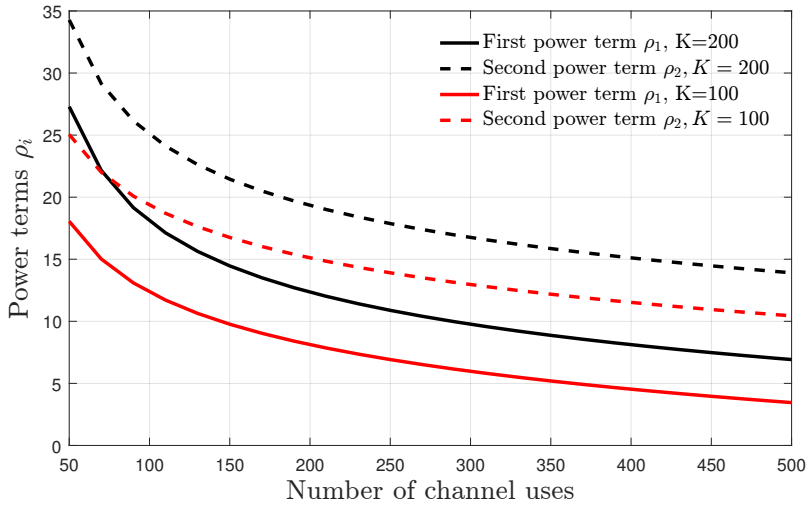


(c) IR-HARQ

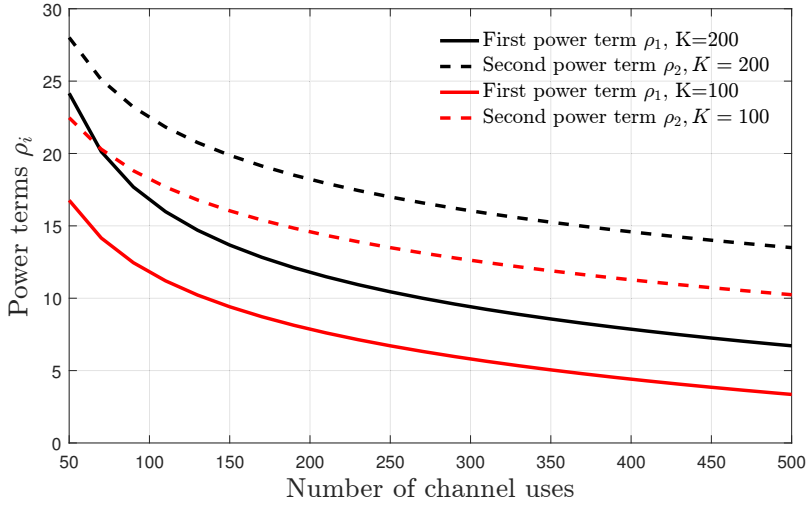
Figure 5.4: Comparison between our proposed power allocation algorithms and the conventional equal power allocation approach for Type-I ARQ given in 5.4a, CC-HARQ given in 5.4b and IR-HARQ protocol given in 5.4c. The channel coding rate is $R = 1$ ncpu, while the maximum number of transmissions is $M = 2$ and $M = 3$.



(a) Type-I ARQ



(b) CC-HARQ



(c) IR-HARQ

Figure 5.5: Power terms different retransmission protocols as a function of the number of channel uses for different number of information nats. They are shown in Figure 5.5a, 5.5b and 5.5cfor Type-I ARQ, CC-HARQ and IR-HARQ protocols.

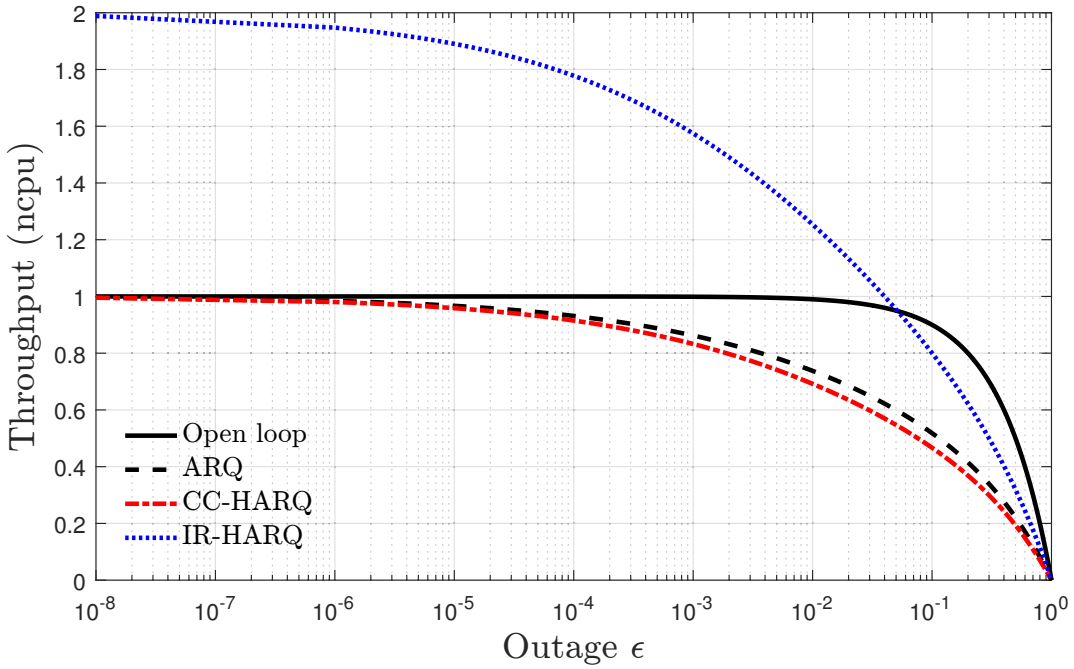


Figure 5.6: Throughput as a function of outage for the three protocols when $R = 1$ ncpu and $M = 2$ transmissions.

5.2. Throughput and delay

The corresponding outage terms for the power allocation scheme are computed by substituting the power terms in (17), (41) and (68) for ARQ, CC-HARQ and IR-HARQ protocols. To calculate the throughput we substitute these outage terms and M in (96). The behavior of the throughput as a function of the target outage probability for different number of transmissions is illustrated in Fig. 5.6. Here, we set $M = 2$ and assume no feedback delay, thus $D = 0$. Further, we map 200 information nats into 200 channel uses. The first observation we make is that the IR-HARQ protocol gives much better performance than the other two protocols. Furthermore, notice that there is a loss in the throughput when repetitive retransmission schemes are utilized. These results are coherent with Figure 4.1, where we analyzed the spectral efficiency. However, we observe that the proposed power allocation scheme helps in mitigating this loss for very low outage values and number of transmissions in the case of repetitive schemes (when $M = 2$ the throughput is almost same as the open loop setup).

Another interesting result from Figure 5.6 is the fact that when $M = 2$ and our power allocation algorithm is implemented, Type-I ARQ protocol exhibits slightly higher throughput than CC-HARQ protocol. This result comes from the well known trade-off between the transmission power and overall system throughput. Since our goal is to save power, we will have to sacrifice throughput. However, the proposed algorithms help in minimizing this loss. Fuerthermore, this result is coherent with what we obtain mathematically. From equations (17) and (41) we notice that the outage probability is inverse proportional to the transmitted power. Moreover, from (96) we notice that the throughput is inverse proportional with the outage probability up to a certain round. Therefore, the lower the power allocated to a certain round, the higher its outage will be. This will then be reflected in lower throughput.

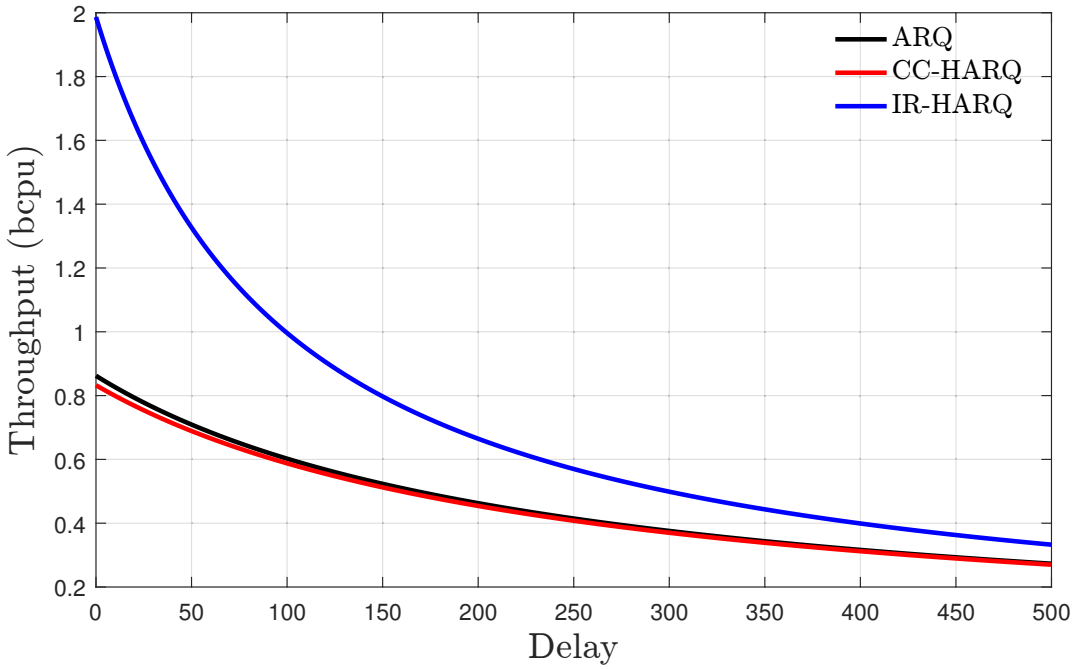


Figure 5.7: Throughput as a function of delay in the ultra reliable region. Here, we fix $\epsilon = 10^{-3}$, $R = 1$ ncpu and $M = 2$ transmissions.

The behavior of the throughput as a function of delay is investigated in Fig. 5.7. The results are shown for coding rate $R = 1$ ncpu and target outage probability $\epsilon = 10^{-3}$. From Fig. 5.7 we observe that the retransmission and feedback delays have a great impact in the throughput for a fixed ϵ despite the protocol that is implemented. However, in ultra reliable systems this loss can be tolerated, especially when considering the larger gains attained through the proposed power allocation scheme, as observed in Section 5.1.

6. CONCLUSIONS AND FUTURE WORK

In this thesis, we focused on systems which have very stringent requirements on latency and reliability. In Chapter 2 we introduced the system model and defined the communication at finite block-length. Furthermore, we analyzed the fastest scheme that can be used to convey the information, the one shot transmission. We showed that operation in the URR would not be feasible under this setup, since very large powers are required with each transmission. For this reason, in Chapter 3 we suggested the utilization of repetitive and parallel retransmission schemes, which can benefit from diversity. Specifically, we analyzed three popular protocols that embrace these schemes, such as type-I ARQ, CC-HARQ and IR-HARQ. In the case of IR-HARQ, we proposeed a closed form approximation for the outage probability, which was later utilized in our derivations. For all three protocols, we proposed globally optimal power allocation algorithms with low complexity, that guarantee operation anywhere in the ultra reliable region while spending minimal power. We showed that despite the protocol which is implemented, the optimal power allocation strategy to operate in the URR suggests transmission with incremental power in each round. Knowing that the implementation of such schemes can affect the throughput and delay of the system, in Chapter 4 we have introduced the necessary analytical framework needed for further analysis. Therein, we analyze the spectral efficiency of the protocols and obtain closed-form expressions for their throughput and average delay. Furthermore, we quantify the average delay associated with a certain number of transmissions. Lastly, in Chapter 5 we provide some numerical results. Therein, we have shown that implementing retransmission protocols can provide large average and maximum power gains with respect to the one shot transmission and the conventional equal power allocation strategy. Furthermore, we have shown that our proposed power allocation algorithms can maximize the overall system throughput as well, with or without the presence of feedback delay. This becomes of vast importance in the case of repetitive schemes, since they generally suffer from large throughput losses.

As future work, we intend to further investigate the trade-offs discussed in Chapter 3 and Chapter 4. Moreover, it would be of high interest to analyze the performance of these protocols in scenarios of cooperative communications with the purpose of increase reliability and as well range extension. We also aim to investigate non-orthogonal transmission schemes associated with retransmission protocols, so to exploit the trade-off between spectral efficiency and reliability.

7. APPENDICES

7.1. Proof that $\rho_m < \rho_{m+1}$

From (36) we can write ρ_m and ρ_{m+1} as:

$$\sqrt{2^{a(m)}\phi^{b(m)}(M\lambda)^{c(m)}2\phi} < \sqrt{2^{a(m+1)}\phi^{b(m+1)}(M\lambda)^{c(m+1)}2\phi}. \tag{104}$$

Next, we make some basic mathematical manipulations

$$2^{a(m)}\phi^{b(m)}(M\lambda)^{c(m)} < 2^{a(m+1)}\phi^{b(m+1)}(M\lambda)^{c(m+1)}, \tag{105}$$

$$\frac{2^{a(m)}\phi^{b(m)}M^{c(m)}}{2^{a(m+1)}\phi^{b(m+1)}M^{c(m+1)}} < \frac{\lambda^{c(m+1)}}{\lambda^{c(m)}}, \tag{106}$$

$$2^{a(m)-a(m+1)}\phi^{b(m)-b(m+1)}M^{c(m)-c(m+1)} < \lambda^{c(m+1)-c(m)}. \tag{107}$$

In (107) the exponent of λ is a positive number. To prove that, we can show that $c(m+1) - c(m) > 0$, which leads to

$$2^{-M+m+1} - 2^{-M+m} > 0. \tag{108}$$

$$2^{-M+m+1} > 2^{-M+m}, \tag{109}$$

$$\log_2 2^{-M+m+1} > \log_2 2^{-M+m}, \tag{110}$$

$$-M + m + 1 > -M + m \tag{111}$$

$$1 > 0 \tag{112}$$

In (109) we make few manipulations and then take logarithm of both sides. After that we apply the property of the exponent of the logarithm ($\log p^q = q \log p$), where p and q are both positive numbers, and $p \neq 1$. Afterwards, can clearly observe that the inequality holds. Next, we substitute the value of λ from (40). We observe that as $\epsilon \to 0$, the right side of the inequality in (107) will tend to infinity. Since all the transformations we have done are equivalent, we can argue that $\rho_m < \rho_{m+1}$.

7.2. Proof that the optimization problem is convex

In this appendix we show the convexity of the optimization problem for Type-I ARQ protocol. Similarily, one can prove the convexity of the problem also for the case the other retransmission protocols.

To prove the convexity of the problem, we will utilize the second order condition, i.e we prove that the Hessian matrix of the objective function is positive definite [48]. Mathematically this is expressed as

$$x^H H x \geqslant 0 \tag{113}$$

where $H \in \mathcal{R}^{M \cdot M}$ denotes the Hessian matrix and $x \in \mathcal{R}^{M}$. The Hessian matrix can be computed as

$$H = \begin{bmatrix} \frac{\partial^2 + rho_{avg}}{\partial \rho_1^2} & \frac{\partial^2 \rho_{avg}}{\partial \rho_1 \partial \rho_2} & \cdots & \frac{\partial^2 \rho_{avg}}{\partial \rho_1 \partial \rho_M} \\ \frac{\partial^2 \rho_{avg}}{\partial \rho_1 \partial \rho_2} & \frac{\partial^2 \rho_{avg}}{\partial \rho_2^2} & \cdots & \frac{\partial^2 \rho_{avg}}{\partial \rho_2 \partial \rho_M} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \rho_{avg}}{\partial \rho_1 \partial \rho_M} & \frac{\partial^2 \ \rho_{avg}}{\partial \rho_2 \partial \rho_M} & \cdots & \frac{\partial^2 \rho_{avg}}{\partial \rho_M^2} \end{bmatrix} \tag{114}$$

As next step, we compute all the partial derivatives. Then, we substitute H in (113) and we can show that the Hessian is positive-semidefinite.

Solving (113) for the general case of M transmissions would result in very complex expressions. Furthermore as discussed above, we can not have very large number of transmissions when an ultra reliable setup is needed (since this would increase the end-to-end delay). Therefore, we focus on the scenario when we have a maximum number of two transmissions, $M = 2$. For this case, the average transmit power is defined in (56). First, we compute the first order partial derivatives.

$$\frac{\partial \rho_{avg}}{\partial \rho_1} = 1 - \frac{\rho_2 \phi}{\rho_1^2}, \tag{115}$$

$$\frac{\partial \rho_{avg}}{\partial \rho_2} = -\frac{\phi}{\rho_1}, \tag{116}$$

Then, we compute the second order derivatives.

$$\frac{\partial^2 \rho_{avg}}{\partial \rho_1^2} = \frac{2\rho_2}{\rho_1^3} \phi(R), \tag{117}$$

$$\frac{\partial^2 \rho_{avg}}{\partial \rho_2^2} = 0, \tag{118}$$

$$\frac{\partial^2 \rho_{avg}}{\partial \rho_1 \partial \rho_2} = -\frac{1}{\rho_1^2} \phi(R). \tag{119}$$

The next step would be to write the Hessian matrix as

$$\begin{aligned} H &= \begin{bmatrix} \frac{\partial^2 \rho_{avg}}{\partial \rho_1^2} & \frac{\partial^2 \rho_{avg}}{\partial \rho_1 \partial \rho_2} \\ \frac{\partial^2 \rho_{avg}}{\partial \rho_1 \partial \rho_2} & \frac{\partial^2 \rho_{avg}}{\partial \rho_1^2} \end{bmatrix}, \\ &= \phi(R) \begin{bmatrix} \frac{2\rho_2}{\rho_1^3} & -\frac{1}{\rho_1^2} \\ -\frac{1}{\rho_1^2} & 0 \end{bmatrix}. \end{aligned} \tag{120}$$

Further, we analyze the product $x^H Hx$

$$x^H Hx = \phi(R) \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \frac{2\rho_2}{\rho_1^3} & -\frac{1}{\rho_1^2} \\ -\frac{1}{\rho_1^2} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \tag{121}$$

$$= \phi(R) \begin{bmatrix} x_1 \left(\frac{2\rho_2}{\rho_1^3} - \frac{1}{\rho_1^2} \right) & -\frac{x_2}{\rho_1^2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \tag{122}$$

$$= \phi(R)(\frac{\rho_2 x_1^2}{\rho_1^3} - \frac{x_1 x_2}{\rho_1^2}), \tag{123}$$

$$= \phi(R)[(\frac{\sqrt{\rho_2} x_1}{\sqrt{\rho_1^3}} - \frac{x_2}{2\sqrt{\rho_1 \rho_2}})^2 - \frac{x_2^2}{4\rho_1 \rho_2}]. \tag{124}$$

From Section 3-A we know that the power terms $\rho_m > 0$. Furthermore, we know from the constraint set that $\frac{\phi^2}{\rho_1 \rho_2} = \epsilon$. Since in URC the goal is to have $\epsilon \to 0$, we can deduct that the negative term in (124) is very small. Thus, (124) reduces to

$$x^H Hx = \phi(R) \left(\frac{\sqrt{\rho_2} x_1}{\sqrt{\rho_1^3}} \right)^2. \tag{125}$$

It is obvious that (125) is non-negative, despite the value of x_1. Thus we can conclude that the optimization problem is convex. Similarly, one can prove that the optimization problem for the case of more transmissions as well.

7.3. IR-HARQ outage probability deriation

In [50] the author shows that

$$\lim_{s \to \infty} s^{m+1} \mathrm{Pr}\left[\mathrm{u_m} + \mathrm{v_m} < \mathrm{R}\right] = \int_0^R g(R-x) f'(x) dx. \tag{126}$$

where $g(R)$ and $f(R)$ are monotone and increasing and integrable functions, $f'(R)$ is integrable and u_m and v_m are independent random variables that satisfy the following conditions

$$\lim_{s \to \infty} s \mathrm{Pr}\left[\mathrm{u_m} < \mathrm{R}\right] = f(R),$$
$$\lim_{s \to \infty} s^m \mathrm{Pr}\left[\mathrm{v_m} < \mathrm{R}\right] = g(R).$$

In (67) we set $u_m = \log(1+\rho_m)$. It is straightforward to show that, when the channel gains are Rayleigh distributed

$$\mathrm{Pr}\left[u_m < R\right] = 1 - e^{\left[\frac{e^R - 1}{\rho_m}\right]}, \tag{127}$$

and

$$\lim_{s \to \infty} s \mathrm{Pr}\left[\mathrm{u_m} < \mathrm{R}\right] = e^R - 1. \tag{128}$$

Letting $g_0(R) = 1$ and $g_1(t) = f(t)e^R - 1$, and recursively applying the theorem we obtain

$$\lim_{s\to\infty} s^2 \mathrm{Pr}\left[\sum_{\mathrm{m}=1}^{2} \mathrm{u_m} < \mathrm{R}\right] = \int_0^R g_1(R-x)f'(x)dx \tag{129}$$

$$= R(e^R - 1) + 1. \tag{130}$$

The expression computed in (130) corresponds to $g_2(R)$. By continuing this, we obtain the recursive integral

$$\lim_{s\to\infty} s^M \mathrm{Pr}\left[\sum_{\mathrm{m}=1}^{\mathrm{M}} \mathrm{u_m} < \mathrm{R}\right] = \int_0^R g_{M-1}(R-x)f'(x)dx \tag{131}$$

$$= g_M(R). \tag{132}$$

The recursive integral in (132) provides each of the terms $g_1(R), g_2(R), \ldots g_M(R)$, which are writen as

$$g_1(R) = e^R - 1, \tag{133}$$

$$g_2(R) = 1 - e^R + Re^R, \tag{134}$$

$$g_3(R) = -1 + e^R - Re^R + \frac{R^2 e^R}{2}. \tag{135}$$

We can observa that the integral converges. After futher manipulations, we find its general term as shown in (136).

$$g_m(R) = (-1)^m \left(1 + \sum_{i=0}^{m} \frac{(-1)^i}{(i-1)!} e^R R^{i-1}\right),\ m = 1, \ldots, M. \tag{136}$$

Finally, by using these results, we can approximate the outage probability as shown in (68).

8. REFERENCES

[1] G. Fettweis and S. Alamouti, "5G: Personal mobile internet beyond what cellular did to telephony," *IEEE Communications Magazine*, pp. 140–145, February 2014.

[2] Qualcomm Technologies, "The evolution of mobile technologies," *Qualcomm Technologies white paper*, June 2014.

[3] Ericsson, "Cellular networks for massive IoT," *Ericsson white paper*, January, 2016.

[4] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proceedings of the IEEE*, Sept 2016.

[5] M. R. Palattella, M. Dohler, A. Grieco, G. Rizzo, J. Torsner, T. Engel, and L. Ladid, "Internet of things in the 5G era: Enablers, architecture, and business models," *IEEE Journal on Selected Areas in Communications*, March 2016.

[6] S. Andreev, O. Galinina, A. Pyattaev, M. Gerasimenko, T. Tirronen, J. Torsner, J. Sachs, M. Dohler, and Y. Koucheryavy, "Understanding the IoT connectivity landscape: a contemporary M2M radio technology roadmap," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 32–40, September 2015.

[7] Nokia Networks, "5G for mission critical communication: Achieve ultra-reliability and virtual zero latency," *Nokia networks white paper*, 2016.

[8] Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews, and E. Yaacoub, "Toward massive machine type cellular communications," *IEEE Wireless Communications*, November 2016.

[9] J. J. Nielsen, G. C. Madueño, N. K. Pratas, R. B. Sørensen, C. Stefanovic, and P. Popovski, "What can wireless cellular technologies do about the upcoming smart metering traffic?" *IEEE Communications Magazine*, 2015.

[10] M. Maier, M. Chowdhury, B. P. Rimal, and D. P. Van, "The tactile internet: vision, recent progress, and open challenges," *IEEE Communications Magazine*, may 2016.

[11] H. Shariatmadari, R. Ratasuk, S. Iraji, A. Laya, T. Taleb, R. Jantti, and A. Ghosh, "Machine-type communications: current status and future perspectives toward 5g systems," *IEEE Communications Magazine*, September 2015.

[12] N. A. J. Osman N. C. Yilmaz, "5G radio access for ultra-reliable and low-latency communications," *Ericsson Research Blog*, Jun 2015. [Online]. Available: https://www.ericsson.com/research-blog/5g/

[13] E. Dosti, M. Shehab, H. Alves, and M. Latva-aho, "Ultra reliable communication via CC-HARQ in finite Block-Length," in *2017 European Conference on Networks and Communications (EuCNC): Physical Layer and Fundamentals (PHY) (EuCNC2017 - PHY)*, Oulu, Finland, Jun. 2017.

[14] M. Shehab, E. Dosti, H. Alves, and M. Latva-aho, "On the effective capacity of MTC networks in the finite blocklength regime," in *2017 European Conference on Networks and Communications (EuCNC): Radio Access Technologies towards 5G (RAT) (EuCNC2017 - RAT)*, Oulu, Finland, Jun. 2017.

[15] P. Popovski, "Ultra-reliable communication in 5G wireless systems," in *1st International Conference on 5G for Ubiquitous Connectivity (5GU)*, Nov 2014.

[16] S. Parkvall, A. Furuskar, and E. Dahlman, "Evolution of lte toward imt-advanced," *IEEE Communications Magazine*, February 2011.

[17] E. G. Ström, P. Popovski, and J. Sachs, "5G ultra-reliable vehicular communication," *CoRR*, 2015. [Online]. Available: http://arxiv.org/abs/1510.01288

[18] B. Singh, Z. Li, O. Tirkkonen, M. A. Uusitalo, and P. Mogensen, "Ultra-reliable communication in a factory environment for 5G wireless networks: Link level and deployment study," in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sept 2016.

[19] N. A. Johansson, Y. P. E. Wang, E. Eriksson, and M. Hessler, "Radio access for ultra-reliable and low-latency 5G communications," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, June 2015, pp. 1184–1189.

[20] O. N. C. Yilmaz, Y. P. E. Wang, N. A. Johansson, N. Brahmi, S. A. Ashraf, and J. Sachs, "Analysis of ultra-reliable and low-latency 5g communication for a factory automation use case," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, June 2015, pp. 1190–1195.

[21] A. Goldsmith, *Wireless Communications*. Cambridge, UK: Cambridge University Press, 2005.

[22] N. Michelusi, P. Popovski, and M. Zorzi, "Optimal cognitive access and packet selection under a primary arq process via chain decoding," *IEEE Transactions on Information Theory*, pp. 7324–7357, Dec 2016.

[23] L. Le and E. Hossain, "An analytical model for arq cooperative diversity in multi-hop wireless networks," *IEEE Transactions on Wireless Communications*, May 2008.

[24] M. Jabi, M. Benjillali, L. Szczecinski, and F. Labeau, "Energy efficiency of adaptive harq," *IEEE Transactions on Communications*, Feb 2016.

[25] Y. Deng, L. Wang, M. Elkashlan, K. J. Kim, and T. Q. Duong, "Generalized selection combining for cognitive relay networks over nakagami- m fading," *IEEE Transactions on Signal Processing*, April 2015.

[26] K. Wang and Z. Ding, "Diversity integration in hybrid-arq with chase combining under partial csi," *IEEE Transactions on Communications*, June 2016.

[27] S. H. Kim, S. J. Lee, and D. K. Sung, "Low-complexity rate selection of harq with chase combining in rayleigh block-fading channels," *IEEE Transactions on Vehicular Technology*, July 2013.

[28] E. Dosti, U. L. Wijewardhana, H. Alves, and M. Latva-aho, "Ultra reliable communication via optimum power allocation for Type-I ARQ in finite Block-Length," in *IEEE ICC 2017 Wireless Communications Symposium (ICC'17 WCS)*, Paris, France, May 2017, pp. 5019–5024.

[29] C. Ji, D. Wang, N. Liu, and X. You, "On power allocation for incremental redundancy hybrid arq," *IEEE Transactions on Wireless Communications*, March 2015.

[30] A. Chelli, E. Zedini, M. S. Alouini, J. R. Barry, and M. Patzold, "Performance and delay analysis of hybrid ARQ with incremental redundancy over double rayleigh fading channels," *IEEE Transactions on Wireless Communications*, Nov 2014.

[31] A. Chelli and M. S. Alouini, "On the performance of hybrid-arq with incremental redundancy and with code combining over relay channels," *IEEE Transactions on Wireless Communications*, August 2013.

[32] "IEEE Standard for Air Interface for Broadband Wireless Access Systems–Amendment 3: Multi-tier Networks, 2015 ," *IEEE 802.16q*.

[33] 3GPP, "3GPP TR 22.862: Technical Specification Group Services and System Aspects; Feasibility Study on New Services and Markets Technology Enablers - Critical Communications;," *Rel 14*, vol. 14.0.0, p. 32, 2016.

[34] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, "Massive Machine-Type Communications in 5G : Physical and MAC-Layer Solutions," *IEEE Communications Magazine*, no. September, pp. 59–65, 2016.

[35] P. Larsson, L. K. Rasmussen, and M. Skoglund, "Analysis of rate optimized throughput for large-scale MIMO-(H)ARQ schemes," in *2014 IEEE Global Communications Conference*, Dec 2014.

[36] T. V. K. Chaitanya and T. Le-Ngoc, "Energy-efficient adaptive power allocation for incremental MIMO systems," *IEEE Transactions on Vehicular Technology*, 2016.

[37] T. V. K. Chaitanya and E. G. Larsson, "Optimal power allocation for hybrid ARQ with chase combining in i.i.d. Rayleigh fading channels," *IEEE Transactions on Communications*, May 2013.

[38] ——, "Outage-optimal power allocation for hybrid ARQ with incremental redundancy," *IEEE Transactions on Wireless Communications*, July 2011.

[39] E. Dosti, M. Shehab, H. Alves, and M. Latva-aho, "On the performance of non-orthogonal multiple access in the finite block-length regime," in *First International Balkan Conference on Communications and Networking 2017 (BalkanCom'17)*, Tirana, Albania, May 2017.

[40] S. H. Kim, D. K. Sung, and T. Le-Ngoc, "Performance analysis of incremental redundancy type hybrid ARQ for finite-length packets in AWGN channel," in *GLOBECOM - IEEE Global Telecommunications Conference*, 2013, pp. 2063–2068.

[41] R. Devassy, G. Durisi, P. Popovski, and E. G. Strom, "Finite-blocklength analysis of the ARQ-protocol throughput over the Gaussian collision channel," *International Symposium on Communications, Control and Signal Processing, ISCCSP Proceedings*, pp. 173–177, 2014.

[42] B. Makki, T. Svensson, and M. Zorzi, "Green communication via Type-I ARQ: Finite block-length analysis," in *2014 IEEE Global Communications Conference, GLOBECOM 2014*, 2015.

[43] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4232–4265, July 2014.

[44] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite block-length regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, May 2010.

[45] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Block-fading channels at finite blocklength," in *Wireless Communication Systems (ISWCS 2013), Proceedings of the Tenth International Symposium on*, Aug 2013.

[46] B. Makki, T. Svensson, and M. Zorzi, "Finite block-length analysis of the incremental redundancy HARQ," *IEEE Wireless Communications Letters*, vol. 3, no. 5, pp. 529–532, 2014.

[47] T. V. K. Chaitanya and T. Le-Ngoc, "Energy-efficient adaptive power allocation for incremental mimo systems," *IEEE Transactions on Vehicular Technology*, pp. 2820–2827, April 2016.

[48] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[49] Qualcomm Technologies, "Making 5G NR a reality," *Qualcomm Technologies white paper*, June 2016.

[50] J. N. Laneman, "Limiting analysis of outage probabilities for diversity schemes in fading channels," in *Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE*, Dec 2003, pp. 1242–1246 vol.3.

[51] L. Szczecinski, C. Correa, and L. Ahumada, "Variable-rate transmission for incremental redundancy hybrid arq," in *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*, Dec 2010.

[52] B. Makki and T. Eriksson, "On the performance of mimo-arq systems with channel state information at the receiver," *IEEE Transactions on Communications*, May 2014.